The Dissertation Committee for Jennifer Andrea McKinney Certifies that this is the approved version of the following dissertation:

Elucidation of Non-B DNA-induced Mutagenesis Mechanisms: DNA Repair Proteins are Required for the Processing of H-DNA and Z-DNA in Eukaryotes

Committee:

Karen Vasquez, Supervisor

John DiGiovanni

Rick Finch

Edward Mills

Sean Kerwin

Elucidation of Non-B DNA-induced Mutagenesis Mechanisms: DNA Repair Proteins are Required for the Processing of H-DNA and Z-DNA in Eukaryotes

by

JENNIFER ANDREA MCKINNEY, B.S.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

AUGUST 2016

Dedication

This dissertation is dedicated to my amazing children; never let anyone try to convince you that you cannot do something. I so much love you.

Acknowledgements

First, I would like to thank my mentor, Dr. Karen Vasquez, for allowing me to join her lab, and being the best mentor a graduate student could ask for. She is a role model both inside and outside the lab. From her, I learned valuable lessons about how to do well planned, hypothesis-driven science (i.e., always include good controls!), how to be a good writer and speaker when communicating science, how to deal with the struggles of having family and a career in science, and how to be a great mentor. Karen kept expectations high, and helped me realize I could achieve them. She helped me grow into a confident professional and realize my potential. She has become more than a mentor, but a great friend. I'm very proud to say that I earned my degree while working under Karen, and I will miss her dearly.

I would like to express my gratitude towards my committee members for their guidance throughout my graduate career. Dr. John DiGiovanni has been supportive of my and my research since the beginning and always served as an omnipresent reminder to work hard, work smart, and have some laughs. Dr. Rick Finch has a unique way of offering constructive criticism, while being the nicest, most supportive person. Dr. Ted Mills was very supportive and willing to help with my project as well as coursework. Dr. Sean Kerwin offered insightful comments to my project that added another level to the depth of our research, and was

always very supportive. I'm very lucky to have had such an amazing committee, and I will never forget them.

I fail to find a word that expresses my appreciation for our lab manager, Laura Christensen. Not only is Laura knowledgeable in every area of the lab, but she actually is the nicest, most patient person on the planet. I've learned many things from Laura in the lab, but her level of patience and willingness to help has inspired me to become more patient as a person and a mom. She makes you feel like you are the most important thing in that moment when you need her. She's amazing, and I'll never achieve that level of patience, but I will continue to strive to. I admire her courage to go on great adventures. She's also surprisingly funny, and I love her sense of humor! I will miss working with her very much.

We have had many lab members and I would like to thank them all. Dr. Junhua Zhao took me under her wing and taught me a great deal about the lab and life in general. Dr. Guliang "Graham" Wang constantly pushed for better science and encouraged my independence, while also exerting patience, as I learned. Dr. Anirban Mukherjee for his loud, never a dull moment, entertainment, pushing me to do my best, and endless input. Dr. Wade Rey for being the calm among the storm and never wavering, not making me feel beneath him just because I was a student, and for very thoughtful input to my research. Dr. Albino Bacolla, Dr. Imee Del Mundo, Joanna Tychowski, Scott Spitser, for eating countless lunches with me and making me laugh, or watching my girls for me while I finished lab work. I would also like to thank my undergraduate research

assistant, Madhushree Zope, who amazes me everyday. She is such a hard worker, very intelligent, and extremely kind, and I have no doubt she will do great things in this world. I feel very lucky to have known such a wonderful group of people (including past and present members I don't mention by name), who helped me every step of the way in my graduate career.

Lastly, I would like to thank my family, though too many to list! My entire family has supported, and encouraged me more than I could have ever asked for or deserved. They provided me with much-needed fun and laughter when I needed a break or felt discouraged. It's amazing what the power of laughter and a hug can do to motivate you to push through the hard times. I feel very proud to be a member of this family and hope I can return the sentiment to each and every one of you someday. I want to thank my daughters, for grounding me at times when I lost site of what was truly important, and motivating me to let go and have fun, and never stop dreaming. And lastly, my husband, who let me unleash my crazy on him, so that no one in the lab knew just how stressed I was, for watching the girls when I worked on the weekends, and countless meals…I love you.

Elucidation of Non-B DNA-induced Mutagenesis Mechanisms: DNA Repair Proteins are Required for the Processing of H-DNA and Z-DNA in Eukaryotes

Jennifer Andrea McKinney, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Karen Vasquez

The vast majority of all cancers result from some form of genetic instability, thus it is important to study the mechanisms involved. The integrity of DNA can be influenced by secondary structure, and DNA can adopt alternative structures that do not conform to the Watson-Crick B-DNA helix (i.e. non-B DNA). To date, >12 different types of non-B DNA structures have been described including H-DNA and Z-DNA, and these structure-forming sequences are abundant in the human genome, occurring 1/3,000 and 1/50,000 base-pairs for Z-DNA and H-DNA, respectively.

Non-B DNA can alter DNA metabolism and contribute to the development of many human diseases. Specific to this project, translocations occurring in the *c-MYC* and *BCL-2* genes, which are shown to contain non-B DNA-forming sequences at translocation breakpoint “hotspots”, are characteristic of certain

leukemias and lymphomas. However, the mechanism(s) involved in this process remains undefined. Previously, we found these structures to be mutagenic in bacteria, human cells, and mice, largely by stimulating the formation of DNA double strand breaks (DSBs). We speculated that the helical distortions produced by non-B DNA may be recognized as “damage” by the cell, eliciting an error-prone repair response, resulting in genomic instability.

Using genetic-based mutation-reporter assays, we have shown for the first time that these structures are mutagenic in yeast. Furthermore, we have identified DNA repair proteins from both the nucleotide excision repair (NER) and mismatch repair (MMR) pathways to be involved in H-DNA and Z-DNA-induced mutagenesis via distinct mechanisms in both yeast and human cells. We further characterized the functions of these proteins using biochemical and molecular biology assays and found that they are enriched at sites of H-DNA and/or Z-DNA, and have cleavage activity at or near the structure.

Taken together, these results suggest that non-B structures are processed in an error-prone fashion via various novel structure-specific repair pathways in which repair proteins from multiple pathways cooperate. The results obtained have enhanced our knowledge of DNA structure-induced genetic instability in disease etiology, and will guide future studies in the development of novel strategies to treat and/or prevent genetic diseases.

Table of Contents

List of Tables xv

List of Figures xvi

List of Abbreviations xix

Chapter 1: Introduction and Background 1

1.1. DNA Structure 2

1.1.1. Watson-Crick B-DNA 2

1.2. Non-B DNA Structures 6

1.2.1. Z-DNA 8

1.2.2. H-DNA 10

1.2.3. Cruciform, Hairpins, and G-Quadruplexes 12

1.3. Biological Relevance of Non-B DNA Structures 13

1.3.1. Z-DNA 13

1.3.2. H-DNA 14

1.3.3. Cruciform, Hairpins, and G-Quadruplexes 16

1.4. Non-B DNA and Genetic Instability 17

1.4.1. Z-DNA 18

1.4.2. H-DNA 22

1.4.3. Cruciform, Hairpins, and G-Quadruplexes 24

1.5. DNA Repair Relative to Non-B DNA 25

1.5.1. Nucleotide Excision Repair (NER) 25

1.5.2. Mismatch Repair (MMR) 29

1.6. Hypothesis and Specific Aims 35

Chapter 2: The Role of DNA Repair Proteins in Z-DNA Induced Mutagenesis In Eukaryotes 39

2.1. Abstract 40

2.2. Introduction 41

2.3. Materials and Methods 48

2.4. Results 63

2.4.1. Z-DNA-forming sequences map to breakpoints in human cancer genomes 63

2.4.2. Z-DNA-induced chromosomal instability in yeast 65

2.4.3. Mutation spectra of spontaneous and Z-DNA-induced mutations in yeast 77

2.4.4. Correlation of mutation frequency and spectra data as determined by mutation index (MI) 84

2.4.5. Rad1(XPF) and Msh3 are associated with Z-DNA-forming sequences in yeast 86

2.4.6. Nucleosome positioning is effected by Z-DNA in yeast 89

2.4.7. Z-DNA-induced genetic instability in wild-type and repair-deficient human cells 91

2.4.8. Association of NER and MMR proteins with Z-DNA-forming sequences in human cells 96

2.4.9. In vitro cleavage of Z-DNA by the NER nuclease complex, ERCC1-XPF 100

2.5. Discussion 108

Chapter 3: The Role of DNA Repair Proteins in H-DNA-induced Mutagenesis in Eukaryotes 113

3.1. Abstract 114

3.2. Introduction 115

3.3. Materials and Methods 119

3.4. Results 129

3.4.1. H-DNA-forming sequences map to breakpoints in human cancer genomes 129

3.4.2. H-DNA-induced genetic instability in Saccharomyces cerevisiae 131

3.4.3. H-DNA-induced genetic instability in human cells 134

3.4.4. H-DNA is a substrate for NER-associated endonucleases ... 138

3.4.5. Rad27(FEN1) inhibits genomic instability caused by H-DNA 142

3.4.6. Distinct mechanisms for replication-dependent and replication-independent H-DNA processing 146

3.5. Discussion 150

3.6. Work in progress related to the H-DNA-induced mutagenesis project 153

3.6.1. The role of MMR proteins in H-DNA-induced genomic instability in yeast 154

3.6.2. The role of MMR proteins in H-DNA-induced genomic instability in yeast 155

Chapter 4: Stable Integration of ADAR1 Over-expression Vector in COS-7 Mammalian Cell Line 158

4.1. Abstract 159

4.2. Introduction 160

4.3. Materials and Methods 167

4.4. Results 171

4.4.1. Development of a mammalian cell line with stably integrated pCMV-ADAR1-p150 over-expression vector 171

4.5. Discussion and Future Directions 175

4.5.1. Immediate goals using the established cell lines described in this chapter 175

4.5.2. Long-term goals using the established cell lines described in this chapter 176

Chapter 5: Future Directions, Summary, and Significance 178

5.1 Future Directions 179

5.2. Immediate Goals 180

5.2.1. Determine the extent to which the Z-DNA-binding protein, ADAR1 has an effect on Z-DNA-mutagenesis 180

5.2.2. Further characterize the relationship between repair proteins found to be involved in Z-DNA-induced mutagenesis 181

5.2.3. Double-knock out experiments in yeast and/or human cells . 182

5.2.4. Determine the extent to which longer repeats capable of forming multiple non-B DNA structures effect mutagenesis in yeast 183

5.2.5. Identify additional genes involved in stimulation or inhibition of non-B DNA-induced mutagenesis. 184

5.3. Ongoing Studies 187

5.3.1. Effect of modifications to non-B DNA structures on genomic instability 187

5.3.2. Determine the effects of non-B DNA structures on metabolism, nutrition, and aging, using an in vivo mouse model 189

5.4. Long-term Goals 192

5.4.1. Continue to investigate non-B DNA-induced mutagenesis in a chromosomal context and using an in vivo mouse model . 192

5.4.2. Non-B DNA and disease-related therapeutics 193

5.5. Summary and Significance 196

References 203

Vita 235

List of Tables

Table 1.1. Related studies from the laboratory of Dr. Karen Vasquez..... 21

Table 2.1. List of oligonucleotides used in this study. 50

Table 2.2. Yeast strains developed in this study. 52

Table 2.3. Mutation frequencies listed as a ratio (Z/C) and with background subtracted (Z-C). 75

Table 2.4. Mutation spectra of wild-type and repair-deficient mutant yeast. 83

Table 3.1. List of oligonucleotides used in this study. 126

Table 3.2. Analysis of H-DNA-induced and spontaneous (B-DNA) mutations in human cells. 137

List of Figures

Figure 1.1. Structure of B-DNA 5
Figure 1.2. Non-B DNA Structures 7
Figure 1.3. Crystal structure of Z-DNA 9
Figure 1.4. Detailed H-DNA Structure 11
Figure 1.5. Nucleotide excision repair pathway 27
Figure 1.6. Mismatch Repair Pathway 31
Figure 1.7. Proposed model for non-B DNA-induced mutagenesis 34
Figure 2.1. Schematic of B-DNA and Z-DNA structures 42
Figure 2.2. Crystal structure of Z-DNA 45
Figure 2.3. Enrichment of Z-DNA forming repeats (ZFRs) at human cancer translocation breakpoints. 64
Figure 2.4. Schematic diagram of the Yeast Artificial Chromosome (YAC) used in this study 66
Figure 2.5. *Kar*-mediated transfer of YAC from donor to recipient cells 67
Figure 2.6. YAC fragility assay 68
Figure 2.7. Z-DNA is mutagenic in wild-type yeast 69
Figure 2.8. Potential outcomes of YAC fragility assay 71
Figure 2.9. Z-DNA-induced mutagenesis in wild-type and repair-deficient yeast cells 74
Figure 2.10. Z-DNA-induced mutagenesis with background subtracted 76
Figure 2.11. 5-FOA is toxic to cells expressing *URA3* gene 78
Figure 2.12. Primers used in this study are specific to the YAC 79

Figure 2.13. Z-DNA-induced mutation spectra in wild-type and repair-deficient yeast 82
Figure 2.14. Compilation of mutation frequency and spectra data as mutation index (MI) 85
Figure 2.15. NER and MMR repair proteins are enriched at Z-DNA-forming regions on YACs in yeast..... 88
Figure 2.16. Z-DNA effects nucleosome positioning in yeast 90
Figure 2.17. Cell lines used for mutagenesis assays in human cells..... 92
Figure 2.18. Z-DNA-induced mutagenesis is decreased in human XPF-deficient and MSH2-deficient cells..... 95
Figure 2.19. MSH2-MSH3 is required for ERCC1-XPF enrichment at Z-DNA-forming regions in human cells. 98
Figure 2.20. Formation of Z-DNA on plasmid substrates as detected by S1 nuclease sensitivity..... 101
Figure 2.21. Analysis of Z-DNA-induced DSBs by LM-PCR 103
Figure 2.22. ERCC1-XPF cleaves near Z-DNA-forming sequences in plasmid DNA 106
Figure 2.23. Model for Z-DNA-induced mutagenesis..... 107
Figure 3.1. DNA structures..... 116
Figure 3.2. H-DNA (triplex)-forming sequences (TFRs) are associated with translocations in human cancer genomes..... 130
Figure 3.3. NER-associated H-DNA-induced genetic instability in yeast cells.. 133
Figure 3.4. NER-associated H-DNA-induced genetic instability in human cells 136
Figure 3.5. NER-associated H-DNA-induced processing 140

Figure 3.6. Circular dichroism (CD) spectrum of the triplex fold-back oligonucleotide MCR2-5' 141
Figure 3.7. FEN1 cleavage and inhibition of H-DNA structure-induced mutagenesis 144
Figure 3.8. SiRNA-mediated depletion of FEN1 in human cells 145
Figure 3.9. Association of endonucleases with H-DNA and effects of replication on H-DNA-induced mutagenesis 148
Figure 3.10. Models of replication-dependent and replication-independent H-DNA-induced genetic instability in eukaryotes 149
Figure 3.11. Preliminary data on the role of MMR proteins in H-DNA-induced mutagenesis in yeast and human cells 157
Figure 4.1. DNA structures 161
Figure 4.2. Adenosine deaminase acting on RNA 1 (ADAR1) 165
Figure 4.3. Verification and establishment of stable mammalian cell line over-expressing FLAG-ADAR1-p150 173

List of Abbreviations

°C	Degrees Celsius
5-FOA	5-Fluoroorotic acid
A	Adenine
A-to-I	Adenosine-to-inosine
ADAM-12	A disintegrin and metalloproteinase-12
ADAR1	Adenosine deaminase acting on RNA 1
ADPKD	Autosomal dominant polycystic kidney disease
ALL	Acute lymphoblastic leukemia
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BCL-2	B-cell lymphoma 2
BLM	Bloom syndrome RecQ like helicase
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
c-MYC	V-myc avian myelocytomatosis viral oncogene homolog
$Can^{R/S}$	Canavanine-resistant/sensitive
CD	Circular dichroism
CDDP	*Cis*-diamminedichloroplatinum(II); aka cisplatin

CEN/ARS	Centromere/autonomously replicating sequence
ChIP	Chromatin immunoprecipitation
COS-7	CV-1 in Origin with SV40 (monkey kidney fibroblast)
COSMIC	Catalogue of Somatic Mutations in Cancer
CPK	Creatine phosphokinase
CSA	Cockayne syndrome A
CSB	Cockayne syndrome B
DAI	DNA-dependent activator of IFN-regulatory factors
DHX9	DEAH-box helicase 9
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
Dnl4	DNA ligase IV (yeast)
DSB	Double-strand break
dsDNA/dsRNA	Double-stranded DNA or RNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ERCC1	Excision repair cross-complementation group 1
EXO1	Exonuclease 1
FANCJ	Fanconi anemia complementation group J

FBS	Fetal bovine serum
FEN1	Flap endonuclease 1
G	Guanine
GG-NER	Global genome nucleotide excision repair
H3	Histone H3
HBP	H-DNA binging protein
HeLa	Human cervical cancer cell line from Henrietta Lacks
ICL	Inter-strand crosslink
IgG	Immunoglobulin G
K^+	Potassium ion
kb	Kilobase
KCL	Potassium chloride
LEU2	Leucine biosynthesis
Lif1	Ligase interacting factor
LM-PCR	Linker-mediated polymerase chain reaction
MBr	Major breakpoint region
MecP2	Methyl-CpG-binding protein 2
MFI	Mutation frequency index
$MgCl_2$	Magnesium chloride
MI	Mutation index
min	Minute

mL	Milliliter
MLH1	MutL Homolog 1
mM	Millimolar
MMEJ	Micro homology-mediated end-joining
MMR	Mismatch repair
Mre11	MRE11 homolog A, double strand break repair nuclease
mRNA	Messenger RNA
MSH2	MutS homolog 2
MSH3	MutS homolog 3
MSH6	MutS homolog 6
MSI	Mutation spectra index
Na^+	Sodium ion
NaCl	Sodium chloride
NEB	New England Biolabs
NER	Nucleotide excision repair
ng	Nanogram
NHEJ	Non-homologous end-joining
nm	Nanometer
NRE	Negative regulatory element
nt	Nucleotide
OD	Optical density

ODcase	Orotidine-5’phosphate decarboxylase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PKD1	Polycystin 1, transient receptor potential channeling interacting
PMS2	PMS1 homolog 2, mismatch repair system component
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene fluoride
Rad1	Radiation sensitive; yeast homolog of human XPF
Rad10	Radiation sensitive; yeast homolog of human ERCC1
Rad14	Radiation sensitive; yeast homolog of human XPA
Rad2	Radiation sensitive; yeast homolog of human XPG
Rad26	Radiation sensitive; yeast homolog of human CSB
Rad27	Radiation sensitive; yeast homolog of human FEN1
Rad4	Radiation sensitive; yeast homolog of human XPC
Rad50	Radiation sensitive
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RNAse A	Ribonuclease A

RPA	Replication protein A
S. cerevisiae	*Saccharomyces cerevisiae*
Sae2	Sporulation in the absence of spo eleven
SC	Short control
SCL	Stem cell leukemia
SDS	Sodium dodecyl sulfate
siRNA	Small interfering RNA
SSA	Single-strand annealing
ssDNA	Single-stranded DNA
T	Thymine
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween® 20
TC-NER	Transcription-coupled nucleotide excision repair
TFIIH	Transcription factor II human
TFO	Triplex-forming oligonucleotide
TFR	Triplex-forming repeat
TGR	Targeted gene replacement
TGX	Tris-glycine-extended
TSS	Transcription start site
URA3	Uracil requiring
UV	Ultraviolet

VEGF	Vascular endothelial growth factor
WCE	Whole cell extract
WRN	Werner syndrome RecQ-like helicase
WT	Wild-type
XP	Xeroderma pigmentosum
XPA	Xeroderma pigmentosum group A
XPB	Xeroderma pigmentosum group B (ERCC3)
XPC	Xeroderma pigmentosum group C
XPD	Xeroderma pigmentosum group D (ERCC2)
XPF	Xeroderma pigmentosum group F (ERCC4)
XPG	Xeroderma pigmentosum group A
XRCC4	X-ray repair cross-complementing protein 4
YAC	Yeast artificial chromosome
ZBD	Z-DNA-binding domain
ZBP	Z-DNA-binding protein
ZFR	Z-DNA-forming repeat
Δ	Mutant
μg	Microgram
μL	Microliter
μM	Micromolar

Chapter 1: Introduction and Background

1.1. DNA Structure

The structure of DNA was first described in 1953 by James Watson and Francis Crick (1) along with Rosalind Franklin, Raymond Gosling, and Maurice Wilkins (2-5). This pioneering work, for which a Nobel Prize was awarded in 1962, laid the groundwork that inspired decades of research devoted to studying DNA, the biological macromolecule that holds all information required for life, and is still an active topic of research. Significant advances have been made in our understanding the importance of DNA as a result of the last 50 years of research dedicated to the structure and function of this essential molecule, and it is clear that the structure of DNA is equally important as the genetic information embedded in the sequence.

1.1.1. Watson-Crick B-DNA

The well-known double-helical structure described in 1953 [and again in 1954 in greater detail (6)], termed “B-DNA” consists of two antiparallel strands of DNA (Figure 1.1a). Each strand is made up of four repeating nucleic acids (bases) connected by a negatively charged sugar-phosphate backbone. The individual units consist of a purine base [adenine (A) and guanine (G)] or pyrimidine base [thymine (T) and cytosine (C)], and one strand of DNA is hydrogen bonded to the other by specific base pairing (i.e., A pairs with T, and C

pairs with G) (Figure 1.1b) (7). This repeating sequence of 3 billion base-pairs in each human cell makes up our “genetic code”, and contains all the information necessary for a cell to survive, and thus the sequence is extremely important. However, the sequence leads to a *structure* of DNA that allows for the integrity and function of the molecule, establishing the importance of the structure to allow proper function of DNA.

The specific structure, including the structure of base-pairs, allows for very important functions of the DNA molecule. First, the four bases have structures that allow for specific binding to each other: A to T and C to G through hydrogen bonding, as well as binding to the sugar-phosphate backbone of the DNA strands (Fig. 1.1b) (7). Second, the specific sequences as well as the helical shape allow proteins to bind and interact with DNA by recognizing the sequences/structures via exocyclic groups in the major and minor grooves in the helix (Fig. 1.1c) (8). Third, in order to store the nearly 2 meters of DNA into the nucleus of the each cell, the DNA helical structure and negatively charged phosphate backbone allows the molecule to wrap around positively charged proteins to form nucleosomes, which compacts the DNA into organized chromosomes (Fig. 1.1d) (9). Lastly, the double-stranded helix with the negatively charged backbone allows for the bases to remain in the center of the helix, and exclude molecules such as water, which maintains the bonding between base-pairs and the integrity of the DNA molecule (Fig. 1.1e) (8). This also protects the bases from various sources of damage that could cause mutations and genomic instability. These

structural features allow DNA to carry out functions such as gene expression through transcription, efficient replication of the DNA to ensure proper passage of genetic information during cell division, and repair of the DNA if the cell senses damage in the form of variation in the canonical structure or sequence.

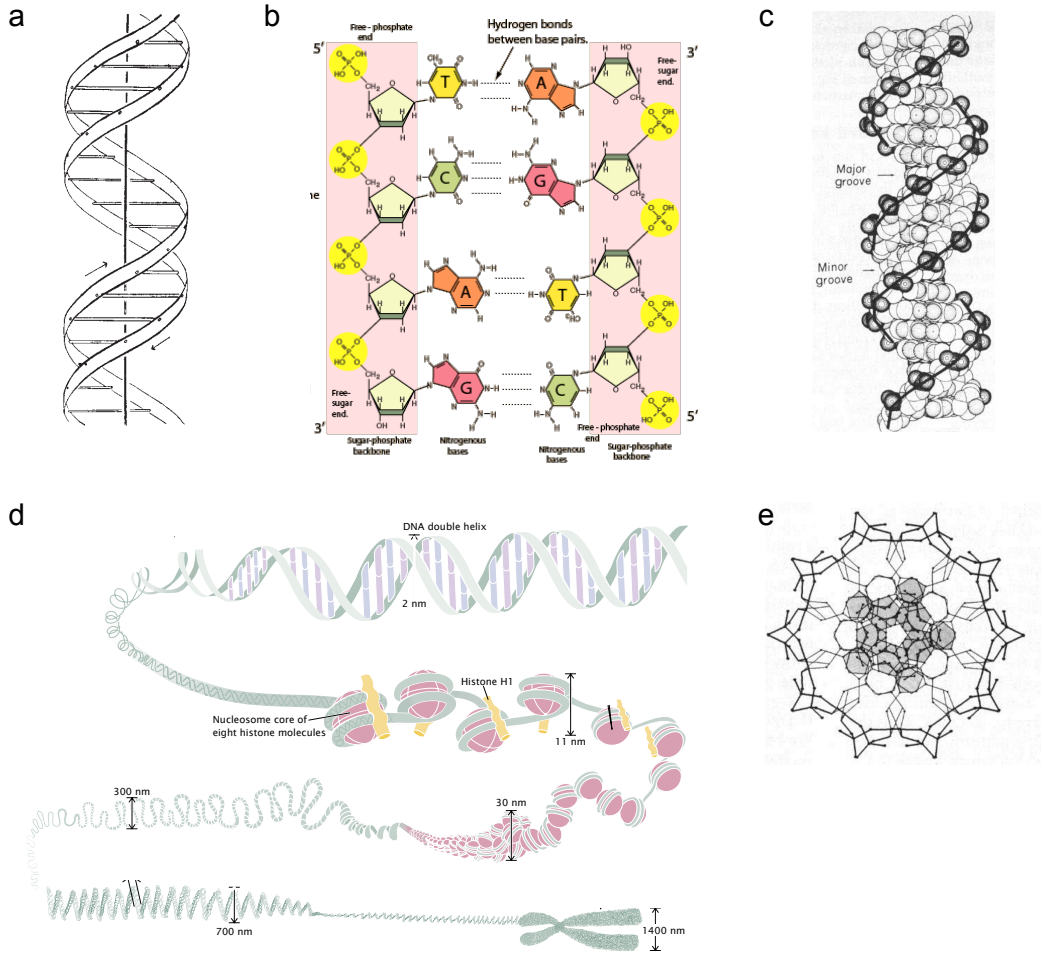


Figure 1.1. Structure of B-DNA. Several aspects of the structure of DNA allow for the function of this important biomolecule. **(a)** The Watson-Crick B-DNA double helix consists of two antiparallel strands [adapted from (1). **(b)** Specific base-pairing between nucleic acids within the DNA strands (A to T and C to G) occur via hydrogen bonding and allow for the maintenance of the genetic code. The base-pairs are anchored by a negatively charged sugar-phosphate backbone [adapted from (7)]. **(c)** The helical turns in the double helix result in the formation of distinctive major and minor grooves that serve as recognition sites for many proteins [adapted from (8)]. **(d)** The negatively charged backbone allows for wrapping of the DNA around positively charged histone proteins to form nucleosomes. This creates supercoiling and allows for the DNA to become compacted and stored as chromosomes within the nucleus of each cell [adapted from (9)]. **(e)** The structure of the DNA double helix allows for the important sequence of bases to remain in the center of the helix while the backbone acts to protect the center and exclude damaging molecules, as seen from an aerial view of the helix [adapted from (8)].

1.2. Non-B DNA Structures

In cells, the right-handed double helical structure described above, known as B-DNA is the most common conformation of genomic DNA. However, DNA itself is a dynamic molecule that can adopt many alternative conformations. Since the late 1950s, over 12 alternative DNA structures have been discovered that do not conform to the canonical B-DNA structure (10-12). Examples include hairpins/cruciform (13-15), G-quadruplexes (16,17), Z-DNA (8,18), and H-DNA (19-21) (Fig. 1.2). Collectively, these alternative structures are called “non B-DNA”. During various metabolic processes (i.e., replication, transcription, and repair), the DNA duplex is opened, resulting in negative supercoiling stress and a high-energy state, which can stimulate the formation of an alternative helical structure at structure-specific repetitive sequences. Nearly half of the human genome is comprised of repetitive sequences, while simple repeat sequences [i.e., those capable of forming several types of non-B DNA structures (22)], account for ~3% of the human genome (23). Thus, non-B-DNA-forming sequences are very abundant in the human genome and warrant the extensive studies that have been performed to determine the physiological role(s) of such sequences/structures. A brief background of cruciform, hairpin, and G-quadruplex structures is included in this chapter; however, the main focus of the studies presented in this document concern H-DNA and Z-DNA, which will be discussed in detail below.

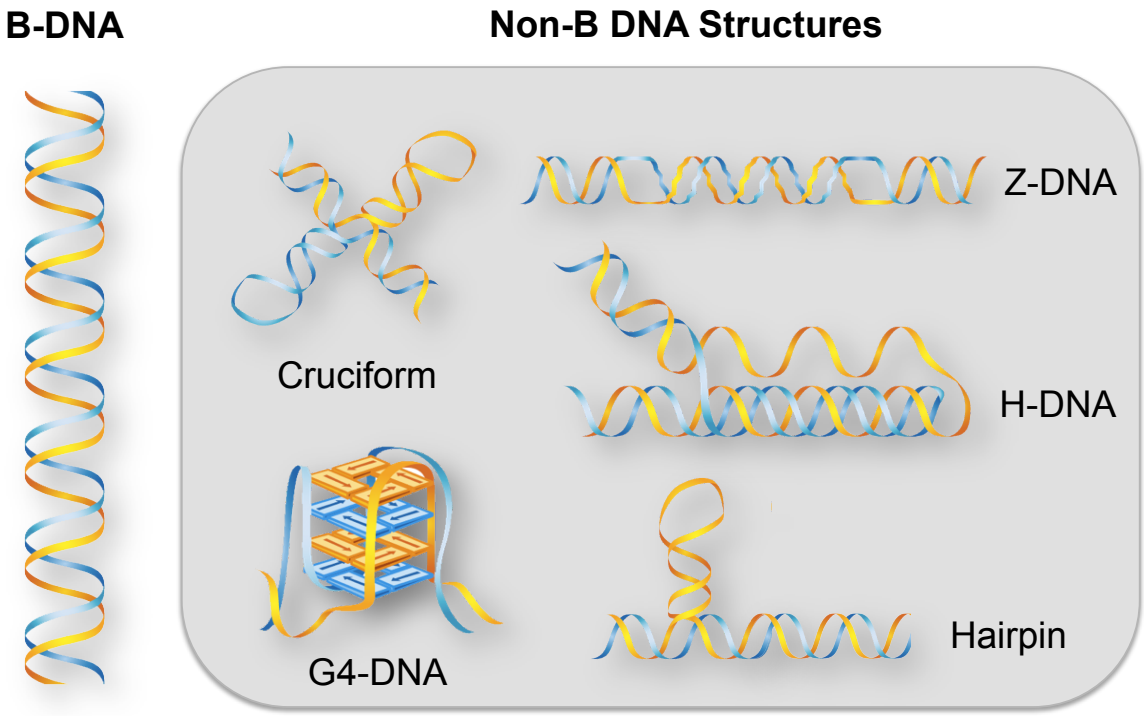


Figure 1.2. Non-B DNA Structures. Watson-Crick B-DNA (left panel) compared to several non-B DNA structures (right panel). Depicted here are cruciform, Z-DNA, H-DNA, G4 DNA (G-quadruplex), and hairpin structures [adapted from (24)].

1.2.1. Z-DNA

Z-DNA is a left-handed, double helix that forms at sequences of alternating purine-pyrimidine repeats (e.g., TG or GC) (Fig. 1.2), and the crystal structure of the left-handed conformation was determined in 1979 (8). While the pyrimidines maintain an *anti* conformation, the purines within the repeat adopt a *syn* conformation in which the base is positioned over the sugar to maintain Watson-Crick base-pairing. The rotation of the purine base that results gives the helix a zigzag appearance, for which the structure was named. The crystal structure also revealed the extrusion of a base-pair present at the B-Z junction, resulting in small, single-stranded regions. Energy from negative supercoiling as a result of an open duplex during DNA processing is required for Z-DNA formation (8,25). The formation of Z-DNA causes several distortions in the helix that differ significantly from the canonical B-DNA conformation. For example, there are more base-pairs per helical turn in Z-DNA (12 bp versus 10.5 bp in B-DNA), a larger distance between bases on the same strand, a smaller diameter than B-DNA, and the distinction between the major and minor grooves in Z-DNA is lost (Fig. 1.3) (8,26,27).

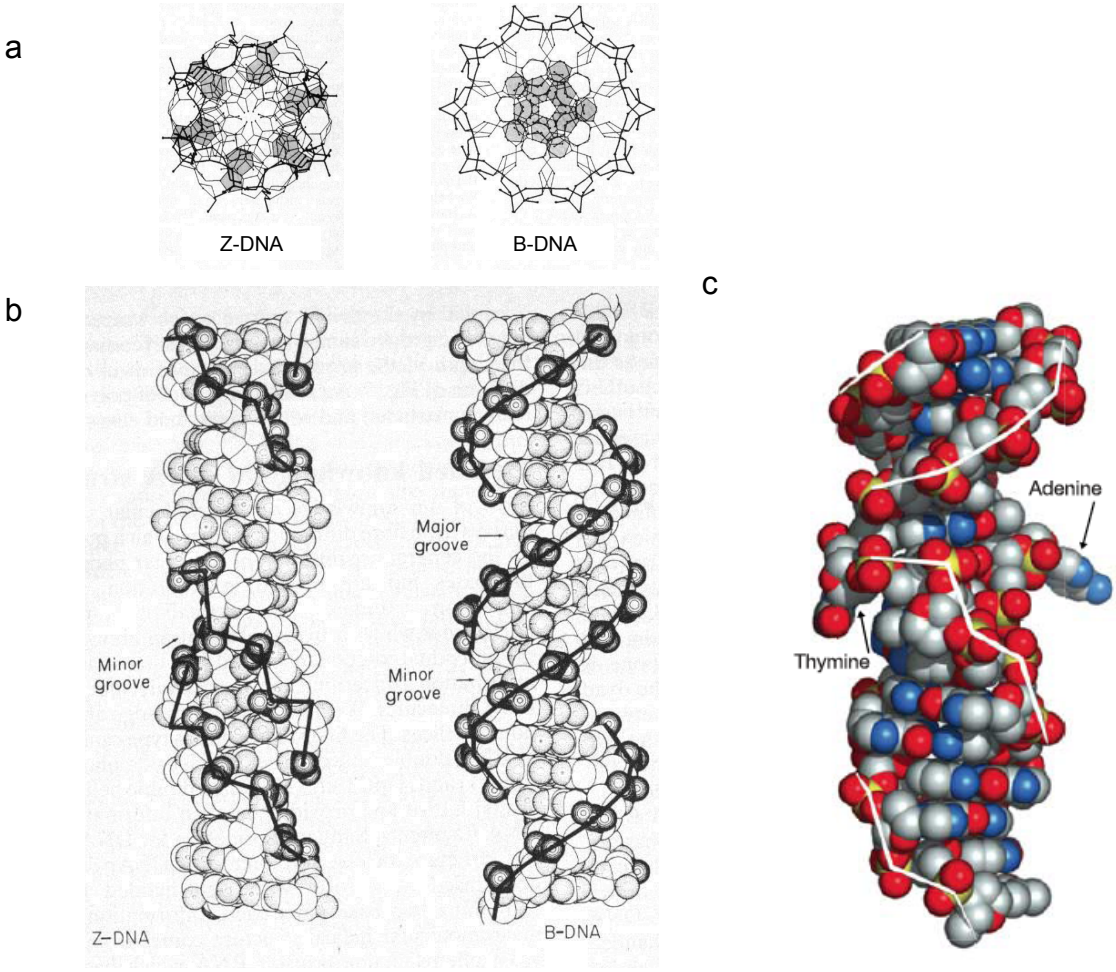


Figure 1.3. Crystal structure of Z-DNA. The resolution of the crystallized structure of Z-DNA revealed significant helical variations from B-DNA. **(a)** An areal view of the two helices demonstrates a smaller diameter in Z-DNA, and the bases are no longer in the protected center of the helix. **(b)** A side-view demonstrates the loss of distinct major and minor grooves in Z-DNA [a and b adapted from (8)]. **(c)** An updated crystal structure confirmed the original and revealed extruded bases (i.e., adenine and thymine) at the B-Z junctions, resulting in a small singled-stranded region [adapted from (18)].

1.2.2. H-DNA

H-DNA (Fig. 1.2) is an intramolecular triplex that forms at mirror-repeats of polypurine-polypyrimidine sequences, occurring when a single-stranded region of DNA (resulting from an open duplex during DNA processing) folds back onto a nearby B-DNA double helix forming a triplex structure. The triplex structure earned the name H-DNA due to the protonation of the cytosine in the third strand that must occur to interact with the duplex. The third strand binds in the major groove of the double helix and is stabilized via Hoogsteen hydrogen bonds (Fig. 1.4) (28,29). Hoogsteen hydrogen bonds have a similar energy to Watson-Crick hydrogen bonds, and can be just as stable (30). H-DNA can exist in different orientations, and can be classified depending on such orientation, including the composition of the third strand. For example, if the third strand is pyrimidine-rich, it will bind to the duplex in a parallel orientation via Hoogsteen hydrogen bonds (Fig. 1.4b); however, a purine-rich third strand will bind in an anti-parallel fashion to the duplex via reverse-Hoogsteen hydrogen bonds, the latter of which can form at physiological pH (20,31).

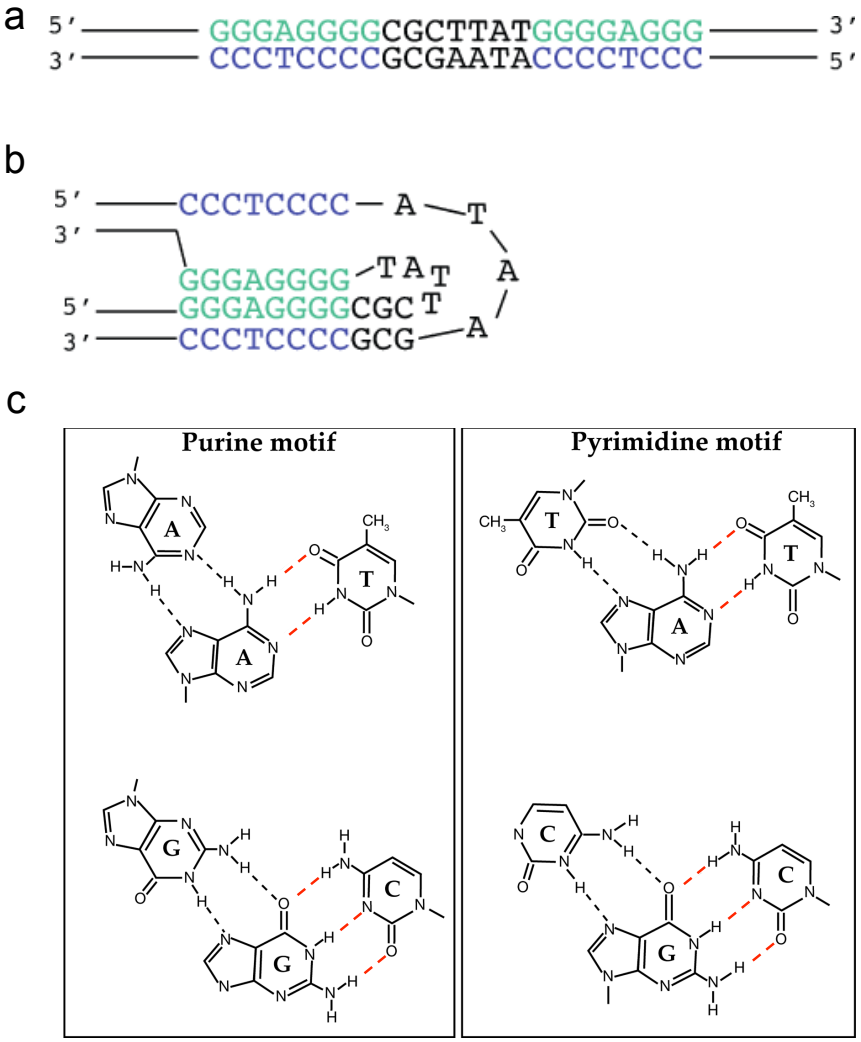


Figure 1.4. Detailed H-DNA Structure. (a) H-DNA (intramolecular triplexes) can form at homopurine:homopyrimidine sequences with mirror symmetry (purine-rich strand in green; pyrimidine-rich strand in blue). **(b)** One motif that H-DNA can adopt is when the 3' end of the purine-rich strand binds to the duplex of the other half of the mirror repeat and is stabilized via reverse-Hoogsteen hydrogen bonds in an anti-parallel orientation [a and b adapted from (32)]. **(c)** Schematic representation of base triplets present in H-DNA motifs (Watson-Crick base-pairing represented by black dotted lines; Hoogsteen hydrogen bonds represented by red dashed lines) [adapted from (31)].

1.2.3. Cruciform, Hairpins, and G-Quadruplexes

Sequences that contain a palindromic region (inverted repeat that reads the same in both directions) are capable of forming hairpin and cruciform structures (i.e., 5'-GACTGC….GCAGTC-3') (Fig. 1.2), resulting from the negative supercoiling that is generated when the DNA duplex is opened (33-35). When the duplex is opened, half of the single-stranded symmetry base pairs with itself to form an intra-strand hairpin stem, leaving the center sequence between the inverted repeat as a single-stranded loop. Cruciform structures consist of two hairpins forming a four-way junction, similar to a Holliday junction during recombination (36), thus the two structures can form at similar sequences. Inverted repeat sequences as short as 14 bp (7 bp/repeat) were shown to be sufficient to form hairpin structures (37), while longer inverted repeats are capable of forming larger, more complex cruciform/hairpin structures (38).

G-quadruplexes (G4 DNA) form at stretches of guanine-rich sequences in which 4 guanines interact via Hoogsteen hydrogen bonds, resulting in a planar structure (Fig. 1.2). At physiological temperature and pH, G4-DNA is stabilized by monovalent cations, K^+ or Na^+, and may be formed by the interaction of one, two, or four G-rich strands, forming various conformations, depending on the sequence, polarity and location of the strands, and microenvironment surrounding the sequence (39).

1.3. Biological Relevance of Non-B DNA Structures

As stated above, nearly 50% of the total human genome is made up of repetitive sequences (3% include simple repeats), and those sequences with the capacity to form non-B DNA structures were found to occur in the genomes of various organisms, though more often in eukaryotic versus prokaryotic genomes (40). The conservation of non-B DNA-forming sequences across multiple species suggests a biological importance of such sequences, which has been extensively studied, as outlined below.

1.3.1. Z-DNA

While the first description of the left-handed Z-DNA structure was in the late 1970's (8), decades of research has dedicated to defining the physiological role of this alternative structure. Z-DNA can form at alternating purine-pyrimidine repeats, and although prokaryotic genomes are more CG-rich than human genomes, Z-DNA-forming repeats are very abundant in the human genome, occurring ~1/3000 bp, as determined by a computer-based search of the human genome (41). In fact, GT repeats are the most common simple repeat in the human genome, accounting for ~0.25% of the entire genome (23).

The distribution of Z-DNA-forming regions was not found to be random, but rather concentrated near transcription start sites (TSS), particularly in

eukaryotes (41,42). Additionally, Z-DNA-specific antibodies were shown to bind primarily in actively transcribed regions, suggesting Z-DNA formation can be induced by transcription (43-47). In line with the requirement of energy from negative supercoiling (that occurs during transcription) for the formation of Z-DNA, and the placement of Z-DNA-forming structures at TSS, Z-DNA has been widely implicated to play a role in gene expression through regulation of transcriptional activity (48-50). Interestingly, Z-DNA has been shown to either promote or suppress transcriptional activity via various mechanisms including stabilizing open chromatin structure, and blocking RNA polymerases (51-60).

The discovery of proteins that contain a structure-specific Z-DNA binding motif (61) across multiple species solidified the theory that Z-DNA is biologically relevant. A number of Z-DNA-binding proteins (ZBPs) have been identified in virus (62-65), bacteria (66,67), yeast (68), Drosophila (69,70), zebrafish (71), mammalian (72-74), and human cells (67,74-76). While the functions of ZBPs are still not fully elucidated, studies have suggested a role in gene expression, recombination, RNA editing, and viral pathogenicity [reviewed in (77,78)]. Human ZDP, ADAR1 is described in detail in Chapter 4.

1.3.2. H-DNA

The first polynucleotide triple-helix was described in 1957 (79), and significant advances in the field of H-DNA-associated research has enhanced our

knowledge of the physiological function of this alternative structure. Computer-generated analyses across genomes from multiple sequences revealed H-DNA-forming sequences to be unexpectedly more abundant in eukaryotes than prokaryotes, occurring ~1/50,000 bp in human cells (80,81). A search for long polypurine-polypyrimidine tracts (>100 bp) in humans found concentrated areas in introns, promoter regions, and untranslated regions of genes associated with cell membrane, phosphorylation, signal transduction, and development and morphogenesis (82). The frequency that which H-DNA-sequences are found in promoter regions is higher than that of random distribution, suggesting H-DNA may play a role in gene expression (83). We, and others have demonstrated the effect of H-DNA on up- or down-regulating gene expression, including evidence that H-DNA causes transcription pausing (84-96), while other reports found no effect on gene expression from H-DNA (97,98). This discrepancy is likely due to multiple mechanisms behind H-DNA-regulation of gene expression, which have yet to be fully elucidated. Several H-DNA-binding proteins (HBPs) have also been described, mostly specific for the purine- or pyrimidine-rich single-stranded region of H-DNA (99-107), further implicating H-DNA in gene regulation and other cellular processes.

1.3.3. Cruciform, Hairpins, and G-Quadruplexes

The biological function of cruciform and hairpin structures has been widely studied by many research groups, and has been found to play a role in several DNA metabolic processes. Analysis of eukaryotic and prokaryotic genomes revealed that inverted repeats capable of forming cruciform/hairpin structures have been shown to occur ~1/40,000 bp in the human genome and were more abundant in human and yeast genomes than in *E. coli* (42). Inverted repeats have been shown to occur in replication origins of both bacteria and viruses, and be required for DNA replication in both eukaryotes and prokaryotes (108,109). Additionally, mRNAs that form hairpins can affect the speed of transcription (110). Moreover, inverted repeats are largely abundant in the euchromatic region of the central male-specific segment on the human Y chromosome (111), and it has been suggested that cruciform structures in this region play a role in evolutionary conservation of the sex-specific functions of the Y chromosome (24).

Analyses of G4 DNA-forming sequences revealed nearly 37,000 sites in the human genome (~1/10,000 bp), the majority in noncoding regions (112,113). G-quadruplex motifs are frequently found in telomeres (114,115) and promoter regions (often at transcription start sites) (116,117), but can also occur in mitochondrial DNA and immunoglobulin switch regions (16). Furthermore, longer repeats capable of forming G4 DNA were unable to form nucleosome structures

(118) and have often co-localized to nucleosome-depleted regions in human cells (119). Such studies have implicated G4 DNA in telomere maintenance, transcription, and chromatin structure.

1.4. Non-B DNA and Genetic Instability

While several non-B DNA sequences/structures have been shown to play physiologically significant roles in various species, a great deal of research published by our laboratory (60,120-133) and others (20,37,38,46,134-150) has demonstrated that these alternative structure-forming sequences, can also contribute to genomic instability associated with a variety of human diseases. Several of these findings are outlined below (studies from our laboratory relevant to this document are summarized in Table 1.1).

The vast majority of cancers are due to some form of genetic instability. One way that the integrity of the DNA can be influenced is through its structure. Thus, it is important to investigate how, why, and when structure-induced genomic instability occurs in opposition to known important biological functions as discussed in the previous section.

1.4.1. Z-DNA

Z-DNA-forming sequences have been shown to stimulate recA-independent plasmid-to-plasmid recombination events in *E. coli* at a higher frequency than background levels (151), and are correlated with recombination hotspots in eukaryotes (152-155), possibly due to Z-DNA-stimulated stabilization of recombination intermediates, with the potential to stimulate DNA double-strand breaks (DSBs) when resolved [reviewed in (11)].

Computer-based algorithms have been developed (156-158), which allow genomic DNA to be screened for sequences that can adopt non-B DNA structures. In particular, in the human genome, Z-DNA-forming sequences were found to co-localize with mutation "hotspots" such as breakpoints and translocations related to diseases such as cancer. For example, the human *BCL-2* gene is mutated in 90% of follicular B cell lymphomas in which a translocation event transfers the *BCL-2* gene to the immunoglobulin heavy chain gene locus, resulting in over-expression of *BCL-2,* seen in non-Hodgkin's lymphomas (159). Z-DNA-forming sequences were found surrounding a common breakpoint for the *BCL-2*-associated translocation (160,161). Similarly, translocations in the human *c-MYC* gene are associated with ~90% of Burkitt's lymphomas (162), and multiple Z-DNA-forming regions have been found in the *c-MYC* P1 promoter and 3' downstream region near these breakpoints (46,163). Translocations in the *SCL* gene are the most common mutation associated with T-cell acute

lymphoblastic leukemia (ALL) occurring in ~30% of patients (164), and Z-DNA-forming sequences were found at a breakage hotspot in this gene as well (147,165). We have also recently shown that Z-DNA and other non-B-forming sequences are enriched at known translocation breakpoints in human cancer genomes (166).

We have shown that sequences with the capacity to adopt Z-DNA have been shown to be mutagenic in bacteria and human cells, though apparently by distinct mechanisms. In bacteria, a Z-DNA-forming sequence induced small deletions, likely due to replication slippage events (167-170), while in mammalian cells, DSBs and large-scale deletions were seen at significantly higher levels than a non-Z-DNA-forming sequence in replication-independent mechanism (170). We demonstrated that the differences across species are likely due to the difference in availability of repair mechanisms, such as micro homology-mediated end-joining (MMEJ) and non-homologous end-joining (NHEJ), which are not available in most prokaryotes, including *E. coli*. When these mechanisms were added to bacteria, the Z-DNA-induced mutation spectra shifted from small to large deletions, similar to the effect seen in human cells (171). Importantly, when integrated in chromosomes in an *in vivo* mouse model, this same Z-DNA-forming sequence stimulated higher levels of genetic instability, with ~7% of the population undergoing large deletions or rearrangement events near the Z-DNA region (128). Thus, there are distinct mechanisms for Z-DNA-induced genomic instability in different species, and it is possible that the mutations in mammalian

cells are a result of error-generating processing by DNA repair components at the sites of alternative DNA structures, outlined in the section 1.5 below.

Table 1.1. Related studies from the laboratory of Dr. Karen Vasquez

Non-B DNA type	Major Finding	Ref.
Z-DNA	Z-DNA is mutagenic in bacteria and human cells via distinct mechanisms, causing large deletions in human cells, and small deletions in bacteria.	(127)
	Demonstrated that non-homologous end-joining (NHEJ) or micro homology-mediated end-joining (MMEJ) may play roles in Z-DNA-induced mutagenesis in human cells.	(171)
H-DNA	Human XPA and RPA cooperate to bind to and recognize a triplex structure, perhaps as “damaged” DNA.	(120)
	H-DNA is mutagenic in mammalian cells, causing large deletions, and coincides with the propensity to form H-DNA. Microhomology at breaks suggested NHEJ and/or MMEJ was involved in repairing DSBs.	(125)
	A naturally occurring H-DNA-forming sequence from the human *c-MYC* gene can block transcription via inhibiting RNA polymerase.	(32)
	Helicase DHX9 recognizes and binds to H-DNA, and protects against H-DNA-induced mutagenesis, suggesting the helicase unwinds H-DNA to protect genomic integrity.	(172)
	The mismatch repair complex, MSH2-MSH3, and nucleotide excision repair protein, XPA, interact to recognize an intermolecular triplex structure.	(130)
Cruciform	Computer analysis revealed inverted repeats align with microinverions that contributed to the evolutionary divergence of chimps from humans.	(173)
	Inverted repeats are enriched at translocation breakpoints in human cancer genomes, cause double-strand breaks and deletions in yeast and human cells, and stimulate structure-specific cleavage via the human XPF protein.	(133)
Multiple	*In vivo* model demonstrating that stably integrated H-DNA and Z-DNA into chromosomes are mutagenic in mice.	(128)
	Translesion synthesis polymerases (Pol η and Pol κ) prevent genomic instability at the human *c-MYC* sequence capable of forming H-DNA, Z-DNA, and G4 DNA. These polymerases are normally down regulated in human cancers, implicating non-B DNA in cancer etiology.	(129)
	H-DNA and Z-DNA, as well as WRN helicase deficiency cause mutations in human cells via distinct pathways.	(132)
	H-DNA, Z-DNA, and G4 DNA-forming sequences were found to be enriched at translocation breakpoints in cancer genomes, further implicating multiple non-B DNA structures in cancer etiology.	(166)

1.4.2. H-DNA

The notion that triplex structures, such as H-DNA are implicated in stimulating genomic instability (i.e., via homologous recombination) is not new (174). For example, the unbound, looped, single-stranded region of the H-DNA structure may serve as an initiator of recombination by pairing with a homologous region, or providing a substrate for homology searching by other loci. Recombination intermediates such as four-way Holliday junctions would form with the potential to create DSBs when resolved (175-178). Moreover, an H-DNA-forming sequence was found in the mouse immunoglobulin c alpha switch region at sites of recombination and splicing (179). Additionally, an H-DNA-forming sequence has been shown to pause or arrest DNA replication in mammalian cells, potentially causing collapsed replication forks and DSBs if processed (180,181), as well as block transcription (32).

Analysis of sequences with the capacity to adopt H-DNA, particularly those consisting of long polypurine-polypyrimidine mirror repeats revealed that these regions were located in genes that had higher frequencies of alternative splicing and chromosomal translocations (82). Similar to Z-DNA, H-DNA-forming sequences often co-localize with chromosomal mutational “hotspots” for breakage and translocations related to human diseases. For example, the human *c-MYC* gene also contains an H-DNA forming region within the promoter, which overlaps with a major translocation breakpoint found in lymphomas and

leukemias (182-189). Additionally, an H-DNA-forming sequence shown to form structure *in vitro* (141), co-localized with a major breakpoint region (MBr) in the *BCL-2* gene associated with follicular lymphomas; furthermore, disruption of the capacity to form H-DNA reduced the frequency of translocations in this region (142). Finally, the *PKD1* gene, which contains a highly mutable region associated with autosomal dominant polycystic kidney disease (ADPKD) (190), contains 23 mirror repeats (191) that have been shown to adopt H-DNA *in vitro* (192), and cause DSBs and large deletions in bacteria (122), implicating the triplex structure in ADPKD etiology.

Using a plasmid-based mutation reporter assay, our previous work has demonstrated that both a naturally occurring H-DNA-forming sequence from the human *c-MYC* gene, and a model H-DNA-forming sequence were mutagenic in mammalian cells, and the majority of mutations were composed of large-scale deletions and/or rearrangements (125). Further analyses revealed that DSB-formation occurred near the H-DNA locus, and that the breaks were likely due to processing by an end-joining mechanism (125). When the H-DNA-forming sequence from the human *c-MYC* gene was integrated into mouse chromosomes, ~8% of the offspring contained genetic instability, further implicating the H-DNA structure in mammalian genomic instability in a chromosomal context in living organisms.

1.4.3. Cruciform, Hairpins, and G-Quadruplexes

Inverted repeats that can form cruciform/hairpin structures have been implicated in generating deletions, gene amplification, and recombination in prokaryotes and eukaryotes (149,150,193-197). AT-rich palindromic sequences have been co-localized to human translocation breakpoints (198-201). Recently, our laboratory has shown that short inverted repeat sequences, able to adopt a cruciform structure, map to translocation “hotspots” in cancer genomes and stimulate DSB formation and deletions in yeast and mammalian cells (133).

Sequences composed of triplet repeats can form hairpin structures, which have been extensively studied due to replication slippage that can occur, resulting in mutations, deletions, and expansions related to more than 30 neurological diseases [extensively reviewed in (137-139,202)]. Examples include Fragile X Syndrome, Huntington’s disease, Friedreich’s Ataxia, Kennedy disease, and Myotonic dystrophy (203-215).

Using the algorithms mentioned above, G4 DNA-forming sequences were also found near known breakpoints in the P1 promoter of the human oncogene *c-MYC* (216,217), as well as mitotic and meiotic breakpoints associated with Fragile X Syndrome in humans (218-220). Furthermore, G4 DNA structures have been shown to cause stalling at replication forks, potentially resulting in breakage and genetic instability (221-223).

1.5. DNA Repair Relative to Non-B DNA

As outlined above, non-B structures cause significant distortions to the B-DNA double helix, and demonstrate mutagenic potential in various model systems, implicating these structures in disease etiology and evolution [also reviewed in (11,24)]. Distortions to the DNA helix (i.e., damage caused by UV, mismatches, chemical crosslinking, oxidation, etc.) are often recognized by repair proteins and signal a subsequent repair mechanism to correct the damage. Thus, the helical alterations that occur in non-B DNA structure formation may be recognized as “damage” and provide a substrate to stimulate processing by repair proteins. Several DNA repair pathways exist to maintain the integrity of the genome, including nucleotide excision repair (NER) and mismatch repair (MMR), which are relevant to the studies presented here, and are outlined below.

1.5.1. Nucleotide Excision Repair (NER)

The proteins involved in the NER pathway (Fig. 1.5) recognize and remove chemical and physical alterations in the helix caused by bulky DNA adducts, such as thymine dimers and 6,4-photoproducts, caused by UV exposure (224-226). There are two sub-pathways of NER: (1) transcription-coupled (TC-NER), which senses damage in actively-transcribed regions of the

genome, and (2) global genomic (GG-NER), which can repair damage throughout the genome and is not dependent on transcription.

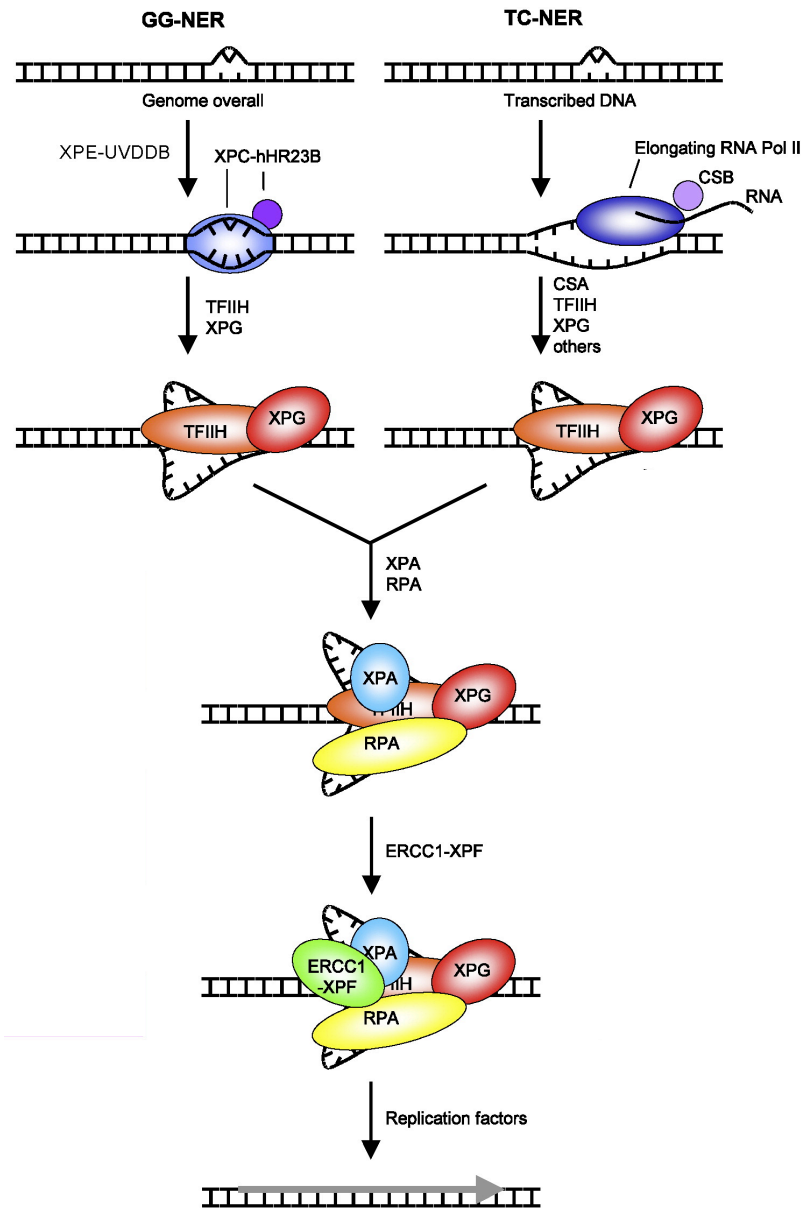


Figure 1.5. Nucleotide excision repair pathway. Pathway responsible for repairing lesions caused by bulky DNA adducts. Two sub-pathways exist in NER: global genomic (GG-NER) and transcription-coupled (TC-NER), which differ only in the initial damage recognition steps. Damage recognition by XPC-hHR23B (GG-NER) or CSB, CSA, etc. (TC-NER) is followed by unwinding of DNA at the site of damage by TFIIH complex, and damaged validation by XPA, while RPA stabilizes single-stranded regions caused by unwinding. Endonucleases ERCC1-XPF and XPG cleave the DNA flanking the damage to excise the region, while replication factors fill and seal the gap [adapted from (226)].

The NER sub-pathways differ in recognition components, while incision, repair, and ligation steps follow a common mechanism. In TC-NER, damage initially signaled by blockage of the transcription process by the damage, which is then recognized by CSB, CSA, and other proteins, which stimulate the repair process. In GG-NER, the primary lesion detector is XPC-hHR23 (now referred to as XPC-RAD23B), which scans the genome for damage to be repaired by the NER pathway. Following the recognition step in both sub-pathways, a sixteen-peptide complex (TFIIH) is recruited to the damage. TFIIH includes XPB and XPD helicases, which unwind the DNA at the damage site. XPA then verifies the damage, while RPA stabilizes and protects single-stranded regions formed during unwinding of DNA. Next a dual incision is made to excise the damage by ERCC1-XPF and XPG structure-specific endonucleases at the 5' and 3' ends of the lesion, respectively, to remove ~30 bp, including the damaged DNA. Finally, replication factors fill in the gap, and ligase seals the nick to complete the repair process (224-226).

As described above, NER proteins recognize and process chemical and physical alterations in the DNA double helix, and it has been demonstrated that the efficiency of the NER pathway to recognize and process DNA damage correlates with the level of distortion (e.g., the more significant the distortion, the more efficient NER is in repairing the damage) (227-230). H-DNA, Z-DNA, and hairpin/cruciform structures significantly deviate from canonical B-DNA helix; therefore, these structures may be recognized as "DNA damage" by NER

recognition proteins, and evidence suggests that non-B DNA may serve as substrates for NER. For example, inter-molecular triplexes between a DNA duplex and a triplex-forming oligonucleotide (TFO) have been shown to induce NER-dependent mutagenesis in mammalian cells (231). We, and others have demonstrated that NER factors are capable of recognizing and processing these intra-molecular triplex structures *in vitro*, and in bacterial and human cells (232-235). Additionally, several reports in human and bacterial cells have shown that NER factors can have opposing effects on hairpin/cruciform structure-induced genomic instability, to either increase or prevent deletions of CAG repeats (236-240) Together, these studies highlight the complexity of non-B DNA recognition and processing, and further studies are warranted to clearly define the role of NER proteins in this mechanism.

1.5.2. Mismatch Repair (MMR)

The mismatch repair pathway (Fig. 1.6) is responsible for correcting base-pairing errors (i.e., mismatches and insertion-deletion loops) that are caused by base deamination, oxidation, methylation, recombination intermediates, and replication errors (241-244). Similar to NER, there are two recognition complexes in MMR depending on the size of the mismatch. The MSH2-MSH6 complex can recognize single-base mismatches or small insertion-deletion loops as damage and signal repair. However, the MSH2-MSH3 complex can also recognize small

loops as well as larger insertion-deletion loops (up to ~10 nucleotides). MLH1 and PMS2 form a heterodimer that is recruited by MSH2-MSH3 (or MSH2-MSH6) complex to facilitate repair, though the exact role of this complex is unknown. In contrast to NER, a single endonuclease, EXO1, is responsible for excising the damaged DNA, while DNA Polymerase δ fills in the gap (241-243).

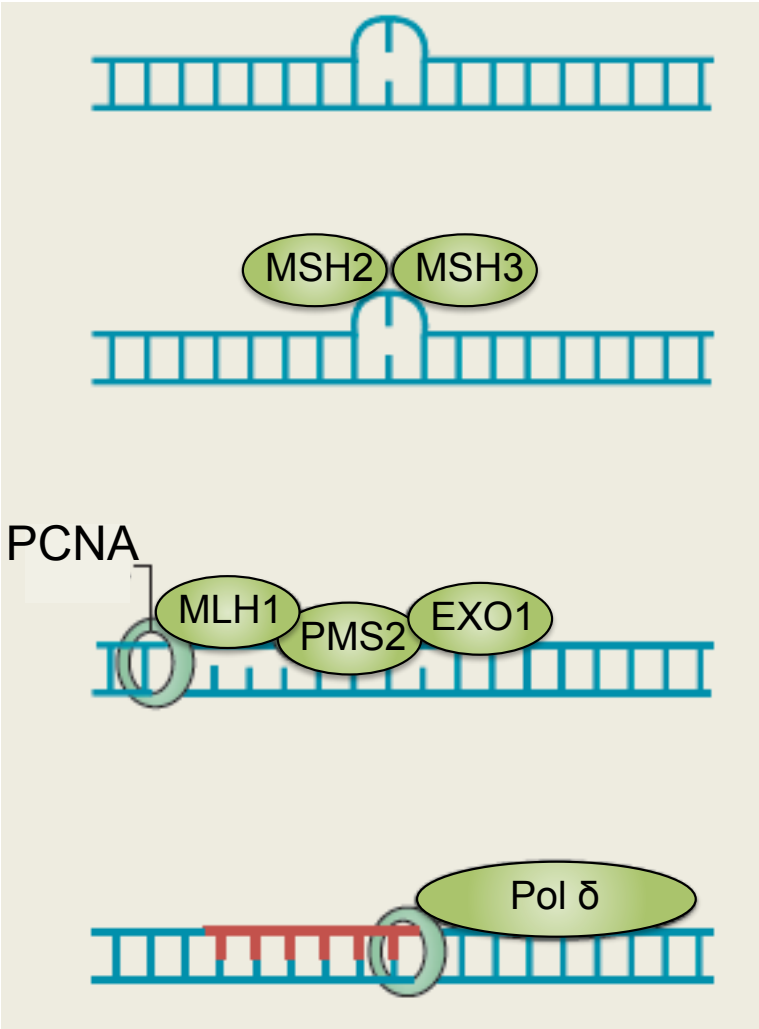


Figure 1.6. Mismatch Repair Pathway. Proteins in the mismatch repair (MMR) pathway are responsible for repairing mismatches including single-base mismatches, and small and large insertion-deletion loops. Damage is recognized by MSH2-MSH3 or MSH2-MSH6 (depicted here) complexes, followed by recruitment of MLH1-PMS2 complex to facilitate repair. EXO1 excises the damage, while DNA Polymerase δ fills in the gap and ligase completes the reaction [adapted from (245)].

Small, single-stranded looped regions can be found on several non-B DNA structures, including hairpin/cruciform conformations and Z-DNA at B-Z junctions (18,246). Therefore, these structures may also serve as substrates for recognition by proteins in the MMR pathway, and several studies have eluded to this theory. For example, in eukaryotic cells MMR proteins have been shown to bind to and/or remove small hairpin/cruciform structures, particularly those containing mismatches, rather than those formed at perfect inverted repeats (37,247,248). Contrasting results were seen comparing mice and humans, in which MMR proteins were found to be responsible for contraction events at CAG repeats in humans (238), and responsible for expansion events of long CNG repeats in mice (249-251), suggesting additional proteins may be involved in CNG repeat processing. Additionally, we have demonstrated that MMR proteins are involved in recognition and processing of TFO-directed inter-molecular triplexes (130,252,253). For example, MSH2-MSH3 can bind to inter-molecular triplex structures *in vivo* and *in vitro*, and can interact with XPA-RPA or XPF-hHR23 to recognize the triplex structure (130).

Taken together, the evidence presented above implicating non-B DNA structures in genomic instability, along with the propensity of non-B DNA structures to cause significant helical distortions of the B-DNA double helix, prompted our laboratory to develop a working hypothesis that repair proteins recognize these distortions as “damage”, and trigger cellular responses for repair, potentially leading to mutations (Fig. 1.7) (254). Studies to determine

which components are involved the mechanism(s) responsible for non-B DNA-related mutagenesis and disease etiology are ongoing, several of which are presented in Chapters 2 and 3.

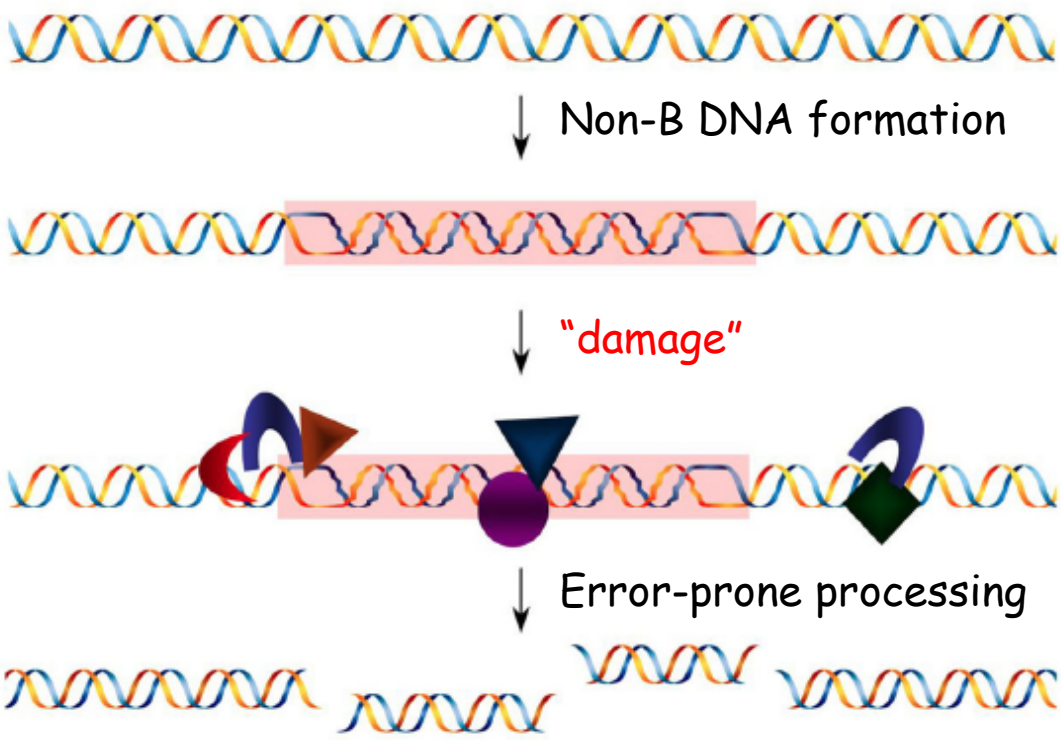


Figure 1.7. Proposed model for non-B DNA-induced mutagenesis. A non-B DNA structure, shown here as Z-DNA, forms as a result of DNA processing (i.e., transcription, replication, repair, etc.). The Z-DNA structure is mistakenly recognized as "damage" by unidentified DNA repair proteins. In an attempt to repair the "damage", subsequent processing of the Z-DNA by additional unidentified repair machinery generates breaks at or near the site of "damage", potentially in an error-generating fashion, resulting in genetic instability [adapted from (254)].

1.6. Hypothesis and Specific Aims

DNA can adopt alternative structures that do not conform to the Watson-Crick B DNA helix (e.g., non-B DNA structures). To date, more than 12 different types of non-B DNA structures have been described including H-DNA and Z-DNA. Sequences that have the capacity to adopt non-B DNA structures are very abundant in the human genome. For example, potential Z-DNA-forming sequences occur approximately once in every 3,000 base pairs, while H-DNA-forming sequences occur ~1/50,000 base pairs. Many non-B DNA-forming sequences have been mapped to regions of common chromosomal breakpoints, known as “hotspots”, which are associated with several human diseases. For example, oncogenic translocations occurring in the *c-MYC* and *BCL2* genes, which are characteristic of cancers such as leukemia and lymphomas, contain Z-DNA-forming sequences at the translocation breakpoints. We have discovered that sequences with the potential to form Z-DNA and H-DNA can stimulate genomic instability in yeast, human cells, and mice, predominantly in the form of DNA double-strand breaks (DSBs). Further, non-B DNA can alter DNA metabolism (i.e., DNA transcription, replication, and repair), which may also contribute to genetic instability. This genomic instability can contribute to the development of human diseases; however, the mechanisms involved in these processes are not well characterized. **We propose the *novel* *hypothesis* that the helical distortions induced by non-B DNA are recognized as “damage”**

and processed by DNA repair machinery via an error-generating mechanism resulting in genomic instability. The *long-term objective* of this project is to determine the mechanism(s) involved in non-B-DNA-induced genetic instability. An *immediate objective* is to test the hypothesis that multiple repair pathways are involved in DNA structure-induced mutagenesis by studying NER and MMR proteins for their involvement in H-DNA- and Z-DNA-induced mutagenesis in eukaryotic cell lines.

The specific aims for the proposed project are as follows:

1. Determine the extent to which NER and MMR proteins impact H-DNA structure-induced mutagenesis in eukaryotes. We have developed a yeast artificial chromosome (YAC)-based genetic instability reporter system containing sequence capable of forming H-DNA or a control insert to screen a yeast knockout library. Using this system, we have shown that our H-DNA-forming sequence is mutagenic, and we will screen several MMR and NER proteins, as well as structure-specific nucleases, for involvement in H-DNA-induced mutagenesis in yeast. Next, we will characterize the mutation spectra in wild-type and mutant yeast strains and compare to the control sequence (i.e. spontaneous background mutants). Critical NER (e.g. XPA) and MMR (e.g. MSH2) proteins, as well as FEN1, will be investigated in available isogenic proficient and deficient human cells, or siRNA techniques will be used to deplete the proteins of interest

in human cells. A plasmid-based *lacZ* mutation-reporter system will be used to measure H-DNA-induced mutagenesis in repair-deficient and wild-type human cells. We will analyze the types of mutations caused by H-DNA and control sequences by direct DNA sequencing of isolated plasmids from mutants.

2. Determine the extent to which NER and MMR proteins impact Z-DNA structure-induced mutagenesis in eukaryotes. Similarly, we will investigate Z-DNA-induced mutagenesis in yeast and human cells using the YAC and plasmid-based mutagenesis assays described in Aim 1 with a $(CG)_{14}$ sequence capable of forming Z-DNA. The same gene products will be screened to determine components involved in this mechanism, and provide evidence as to whether the parallels exists between the mechanisms for genetic instability stimulated by H-DNA or Z-DNA.

3. Characterize the function/pathway or the proteins involved in non-B-DNA-induced mutagenesis. The role of NER and MMR proteins and additional proteins identified in Aims 1 and 2 will be further characterized using *in vitro* assays to determine the mechanism(s) involved in DNA structure-induced genetic instability. Biochemical and molecular biology-based approaches such as Chromatin Immunoprecipitation (ChIP) and Electrophoretic Mobility Shift Assays (EMSA) will be used to investigate the mechanistic role of these proteins in recognizing and/or maintaining these non-B DNA sequences using cell-free extracts and available antibodies and/or purified proteins. To determine the extent to which proteins containing nuclease activity (e.g. XPF) are involved in

processing H-DNA and Z-DNA leading to DSBs, cleavage assays using cell-free extracts and/or available purified proteins and H-DNA-forming oligonucleotides will be performed.

With the results obtained from this project we will gain valuable information regarding the roles of critical gene products and pathways involved in DNA structure-induced genetic instability. This information will enhance our knowledge of how DNA structure influences human disease development, specifically cancer etiology.

Chapter 2: The Role of DNA Repair Proteins in Z-DNA Induced Mutagenesis In Eukaryotes[1]

[1] Sections of this chapter are included in the following publications:

Jennifer A. McKinney, Anirban Mukherjee, Guliang Wang, Junhua Zhao, Albino Bacolla, and Karen M. Vasquez. Cross-talk between DNA repair pathways results in genomic instability at a naturally occurring alternative DNA structure, *Molecular Cell* (submitted) 2016. For this manuscript, I was involved in planning and executing experiments, analysis of data, acquiring partial funding for the project, and preparing and editing the manuscript.

Guliang Wang, **Jennifer A. McKinney**, Albino Bacolla, and Karen M. Vasquez. DNA methylation causes genetic instability by altering DNA and chromatin structures [in preparation]. For this manuscript, I was involved in planning and executing experiments, and analysis of data.

2.1. Abstract

Alternating purine-pyrimidine sequences that have the capacity to adopt alternative Z-DNA structures are abundant in the human genome, with estimates of as many as 1/3000 bp. Importantly, non-B DNA-forming sequences, including Z-DNA often co-localize with mutation hotspots associated with human disease, and we have shown that alternative DNA structures (or non-B DNA) can cause genomic instability, implicating these DNA structures in cancer etiology. However, the mechanisms involved in these processes are not well characterized. Z-DNA causes distortions in the DNA double helix, and therefore, we speculated that these non-B DNA structures may be recognized as “DNA damage” by repair proteins and initiate a DNA damage response. Multiple repair mechanisms may be involved in the processing of non-B structures because they are not typical DNA lesions *per se*, but rather distortions of the B-DNA helix. To identify the proteins/pathways that play a role in Z-DNA structure-induced genetic instability, we screened a yeast knock-out library. We identified the nucleotide excision repair (NER) protein complex, Rad10-Rad1 (XPF-ERCC1), and the mismatch repair (MMR) protein complex Msh2-Msh3, to be involved in Z-DNA-induced mutagenesis in yeast, and confirmed these results in human cells. Using chromatin immunoprecipitation (ChIP) assays, we demonstrated that ERCC1-XPF and MSH2-MSH3 complexes were recruited to Z-DNA-forming sequences and that ERCC1-XPF recruitment was dependent on MSH2-MSH3. Furthermore,

we showed that ERCC1-XPF cleaved DNA near the Z-DNA-forming site in human whole cell extracts. We propose a relationship in which these complexes recognize and attempt to repair Z-DNA in eukaryotic genomes, representing a novel mechanism of genomic instability. Thus, our results in human cells implicate Z-DNA in cancer etiology.

2.2. Introduction

Since the discovery of the B-DNA double helix, by Watson and Crick more than five decades ago (1) (Fig. 2.1a), several non-canonical (i.e., non-B) DNA conformations have been described. Examples include cruciform, H-DNA and Z-DNA, which have been studied extensively by our laboratory (60,120-133) and others (20,37,38,46,134-150). Z-DNA is a left-handed helix that forms at sequences of alternating purine-pyrimidine repeats (e.g., TG or GC) (Fig. 2.1b). The purines within the repeat adopt a *syn* conformation in which the base is positioned over the sugar, giving the helix a zigzag appearance, and results in single-stranded regions at the B-Z junctions. Energy from negative supercoiling is required for the formation of Z-DNA (8,25). Sequences with the ability to adopt Z-DNA are very abundant in the human genome occurring ~1 in 3 kb (41,255).

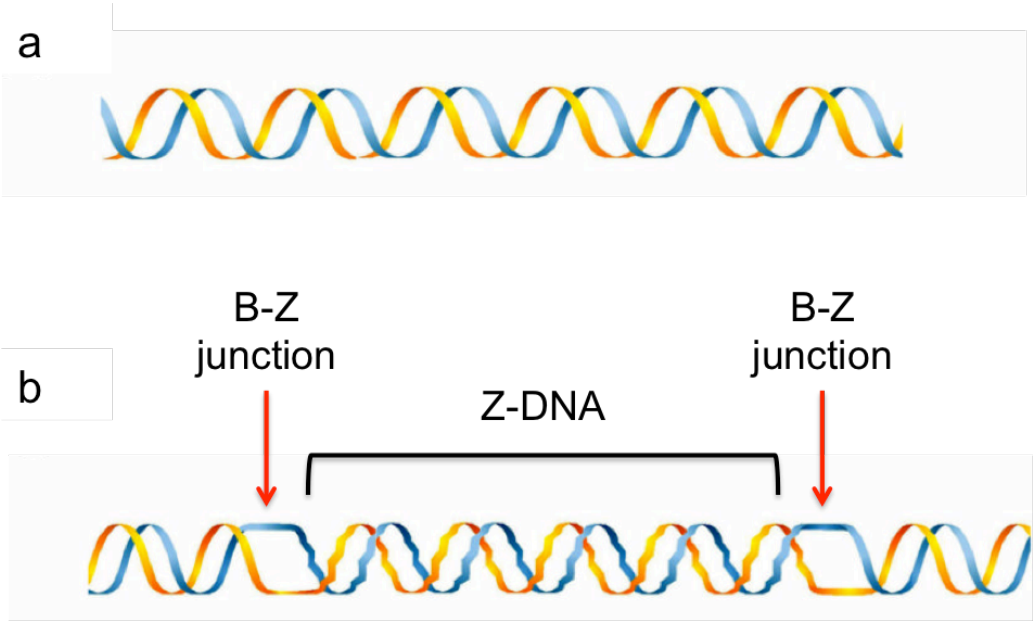


Figure 2.1. Schematic of B-DNA and Z-DNA structures. (a) Watson-Crick double-stranded B-DNA. **(b)** Left-handed Z-DNA structure flanked by two regions of B-DNA. B-Z junctions are labeled (red arrows) [adapted from (256)].

The biological function of non-B-forming sequences has been of interest due to high occurrence and conservation of these sequences across multiple species of eukaryotes (24). While the function of such structures remains to be fully understood, it has been suggested that some of these structures are involved in DNA metabolic processes (11,24). For example, Z-DNA-forming sequences are enriched at transcription initiation start sites, thus it has been proposed that this structure may play a role in regulation of gene expression (257). Z-DNA formation can also alleviate negative supercoiling that occurs during transcription and replication by absorbing the energy formed during the unwinding of B-DNA, further suggesting a role in transcription and replication (18,77).

Non-B DNA-forming sequences can also contribute to genomic instability. Sequences with the potential to adopt Z-DNA map to sites within or near chromosomal hotspots that are associated with translocation breakpoints in human diseases, such as cancer. For example, Z-DNA-forming sequences are located in both the human oncogenes *c-MYC* and *BCL2* near translocation breakpoints associated with leukemias and lymphomas (i.e., Burkitt's Lymphoma) (122,123,161,166,258). Some chromosomal translocations that occur frequently in leukemias and lymphomas result in the *c-MYC* or *BCL2* genes being placed next to the heavy chain immunoglobulin gene causing overexpression of the translocated oncogene (77). We have discovered that Z-DNA-forming sequences are mutagenic in bacteria, mammalian cells and in

mice, and can stimulate the formation of DNA double-strand breaks (DSBs) that may be repaired by non-homologous end-joining (NHEJ) (60,127,128,132,171). These results suggest a role for non-B structures in causing gene translocations through the formation of DSBs, implicating them in cancer etiology; however, the molecular mechanism(s) involved in genomic instability and the generation of the DSBs induced by Z-DNA-forming sequences is not fully understood. Because Z-DNA alters the B-DNA helix there is no longer a canonical major groove within the Z-DNA region, the Z-DNA helix is more compact with 12 bp/turn compared to the 10.5 bp/turn of B-DNA, and the crystal structure revealed that base extrusion is evident at the junctions between B- and Z-DNA (Fig 2.2) (8,18,25). Such differences may be recognized as a form of a DNA “damage” by repair proteins and initiate a DNA damage response.

The nucleotide excision repair (NER) pathway recognizes and repairs helical distortions caused by bulky adducts (224,225), while the mismatch repair (MMR) pathway recognizes and repairs smaller distortions caused by mismatches formed during replication, and small slippage loops (242) . The helical distortions present in Z-DNA could potentially serve as a substrate for recognition and repair by the proteins in the NER mechanism. Furthermore, the B-Z junctions consist of at least one unpaired base, resulting in a small single-stranded region or loop (18,259) that may be a substrate for recognition and repair by the MMR mechanism.

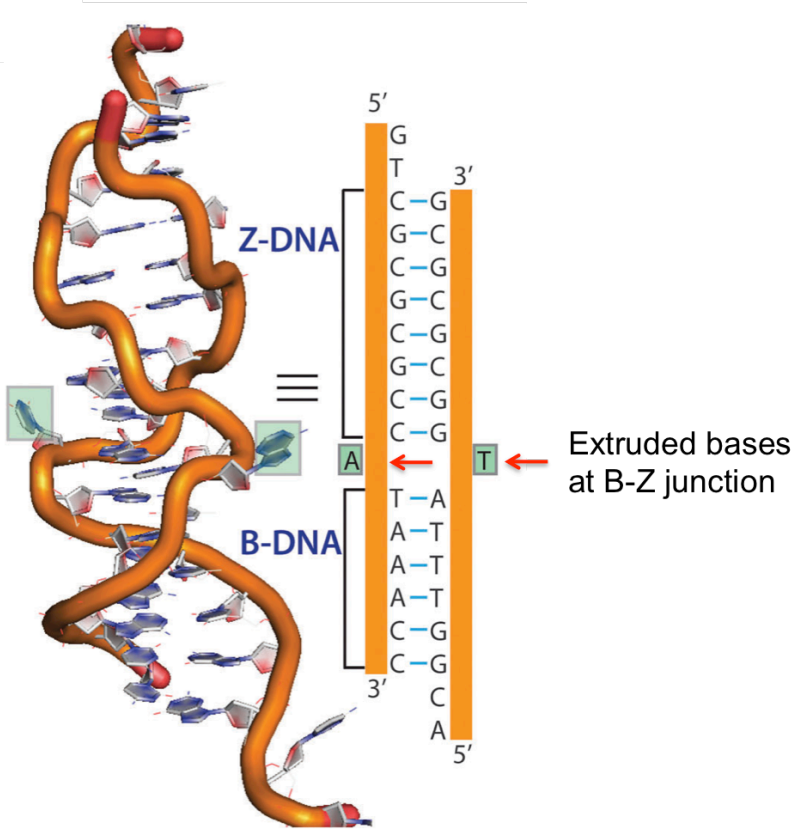


Figure 2.2. Crystal structure of Z-DNA. The crystal structure of Z-DNA (PDB id: 2ACJ) revealed extruding bases at the junctions of B-DNA and Z-DNA (red arrows) [adapted from (12)].

Thus, multiple repair mechanisms may be involved in the recognition and repair of non-B structures because they are not typical DNA lesions *per se*, but rather distortions of the B-DNA helix. Results from our laboratory have shown that NER and MMR proteins are involved in mutagenesis induced by another non-B DNA structure, H-DNA, in yeast and mammalian cells (unpublished data; see Chapter 3 for a details on these findings). Based on these results it is plausible that repair proteins from multiple pathways participate in error-generating repair of Z-DNA structures, resulting in genomic instability. Therefore, we *hypothesize* that NER and MMR proteins are involved in Z-DNA-induced mutagenesis in eukaryotes.

In this study, we investigated the components and mechanisms responsible for Z-DNA-induced mutagenesis. We developed >20 wild-type and mutant yeast strains containing a yeast artificial chromosome (YAC) with either a control B-DNA, or Z-DNA-forming insert flanking a *URA3* mutation-reporter gene. We demonstrated for the first time that a Z-DNA-forming sequence is mutagenic in wild-type yeast, and causes DSBs. By screening mutant strains deficient in various repair proteins, we have identified the NER and MMR repair complexes Rad10-Rad1(ERCC1-XPF) and Msh2-Msh3, respectively, to be involved in Z-DNA-induced mutagenesis in yeast.

We confirmed these results in human cells using a plasmid-based mutation-reporter assay and found that while a Z-DNA-forming sequence is mutagenic in wild-type human cells, this effect is lost in the absence of XPF or MSH2. Furthermore, we demonstrate that both repair complexes are enriched at

the site of Z-DNA-formation and that XPF-ERCC1 enrichment is dependent on MSH2-MSH3 using chromatin immunoprecipitation (ChIP) assays. Finally, we used a plasmid-based *in vitro* primer extension assay and found that ERCC1-XPF can cleave DNA near the site of the Z-DNA-forming insert.

The data presented here suggest a novel relationship between ERCC1-XPF and MSH2-MSH3 outside of NER and MMR to function in Z-DNA-induced mutagenesis. It is clear from these studies that repair proteins from several pathways are involved in Z-DNA-induced mutagenesis in eukaryotes and may contribute to translocation-related disease etiology.

2.3. Materials and Methods

2.3.1. Bioinformatics analyses. We applied an algorithm that we developed to search for Z-DNA-forming sequences (133), to the dataset containing ~20,000 breakpoint locations translocations and deletions mapped to the human cancer genome assembly GRCh37/hg19 obtained from COSMIC at http://cancer.sanger.ac.uk/. Z-DNA-forming repeats (ZFRs) located within ±100 bp from the breakpoints (bins), comprising alternating purine and pyrimidine runs of length ≥6 bases each, were identified using custom scripts. For overlapping ZFRs only the longest match was used for the output. The bedtools utility was used to generate a set of 20,000 non-gap-matching sequences (bins), each 200 bases long (166).

2.3.2. Yeast strains and yeast artificial chromosomes (YACs). The yeast strain 213 cir° (*MATa Kar1-1 leu2-3, 112 ura3-52 his7 cyh2^{R})* was used for YAC construction and transfer into other yeast strains. The yeast strain BY4742 (*MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*) and derivatives (Yeast deletion library GSA-5, ATCC, Manassas, VA) were used in DNA structure-induced YAC fragility and spectra assays. A Z-DNA-forming (CGCGCGCGCGCGCGCGCGCGCGCG) or control (CGAGCTATCTGAGTCGAATACAGTTCGAC) sequence was cloned between the telomere seed G4T4 and the *URA3* gene on a replication-defective plasmid derived from pRS306. The plasmids were then linearized and used to construct

YACs by homologous recombination with YAC VS5 containing a point mutation in the *URA3* gene (260-262). YACs containing Z-DNA-forming and control B-DNA-forming sequences and a functional *URA3* gene (Fig. 2.4) were selected on plates containing synthetic defined (SD) minimal yeast media (Clontech, Mountain View, CA) with complete supplement mixture-uracil (CSM-Ura) (MP Biomedicals, Santa Ana, CA), and confirmed by polymerase chain reaction (PCR) with primers M13-20 and M13Rev (Table 2.1), followed by direct DNA sequencing.

Table 2.1. List of oligonucleotides used in this study.

Region	Name	Sequence
Z-DNA and control on YAC	M13-20	5' TGT AAA ACG ACG GCC AGT 3'
	M13Rev	5' CAG GAA ACA GCT ATG ACC 3'
URA3 gene on YAC	Ura3S1For	5' AGG CGG CAG AAG AAG TAA CA 3'
	Ura3S1Rev	5' AAT GCG TCT CCC TTG TCA TC 3'
LEU2 gene on YAC	Leu2S2For	5' CTG TGG GTG GTC CTA AAT GG 3'
	Leu2S2Rev	5' CCA TCA CCA TCG TCT TCC TT 3'
Z-DNA and control on plasmid	pUinsFor1	5' GTT TTC CCA GTC ACG ACG TT 3'
	pUinsRev1	5' TTT ATG CTT CCG GCT CGT AT 3'
Z-DNA and control on plasmid	548up	5' GGT GAT GAC GGT GAA AAC CT 3'
LM-PCR linker	LMPCR1	5' /5phos/CGT ACA TTC ACA ACG ATA GCG ACT GA 3'
	LMPCR2	5' GCT ATC GTT GTG AAT GTA CG 3'
Z-DNA and control on plasmid	LMlacLeft741	5' ACA GAT GCG TAA GGA GAA AA 3'
Z-DNA and control on plasmid	JMleft	5' GGA GAA AAT ACC GCA TCA GG 3'
	JMright	5' ATT AGG CAC CCC AGG CTT TA 3'
MSH2	MSH2 siRNA	U.C.U.G.C.A.G.A.G.U.G.U.U.G.U.G.C.U.U.U.U.

2.3.3. YAC transfer (*kar*-cross). Donor cells 213 cir° (canavanine-sensitive) containing YACs with the Z-DNA-forming or control B-DNA-forming sequence were grown from a single colony in minimum synthetic defined (SD) base with CSM-Leucine-Uracil media (MP Biomedicals, Santa Ana, CA) overnight at 30°C. A single colony of BY4742 canavanine-resistant recipient cells from the yeast deletion library was grown in yeast extract peptone dextrose media (YPD) at the permissive temperature overnight. YACs were transferred from donor cells to recipient cells via *kar*-mediated transfer as previously described (260,261,263-265). Briefly, equal optical density (OD) of donor cells with YACs and recipient mutant cells were mixed and inoculated into fresh YPD media, grown in a shaking incubator for 6 hours at permissive temperature of 27-30°C, as several strains are temperature-sensitive and must be grown at lower temperatures. The culture was plated on CSM-Ura-Arg+canavanine (100 mg/L) plates, and incubated at permissive temperature. *Kar*-crossed cells were plated for single colonies on CSM-Leu-Ura plates containing canavanine and CSM-lys to confirm transfer (i.e., colonies containing the YAC will grow on CSM-Ura-Leu, but not on CSM-Lys plates) (Fig. 2.5). Cells were analyzed via direct DNA sequencing using the M13-20 primer (flanking the insert) (Table 2.1) to confirm the presence of the YAC and the correct Z-DNA-forming or control B-DNA-forming sequence. See Table 2.2 for a list of strains developed in this study.

Table 2.2. Yeast strains developed in this study.

	Mutant Strain	SC *kar* cross	$(CG)_{14}$ *kar* cross
Wild-type	BY4742	✔ (JZ)	✔ (JZ)
NER	*rad4Δ(XPC)*	✔	✔
	rad26Δ(CSB)	✔	✔
	rad14Δ(XPA)	✔	✔
	rad1Δ(XPF)	✔	✔
	rad2Δ(XPG)	✔	✔
	rad10Δ(ERCC1)	✔	✔
Structure-specific endonucleases	*rad27Δ(FEN1)*	✔	✔
	mre11Δ	✔	✔
	sae2Δ	✔	✔
	rad50Δ	✔ (JZ)	✔
MMR	*msh2Δ*	✔	✔
	msh3Δ	✔	✔
	msh6Δ	✔	✔
	mlh1Δ	✔	✔
	exo1Δ	✔	✔

(SC) short control; $(CG)_{14}$ Z-DNA; (JZ) strains developed by Dr. Junhua Zhao. All other strains developed by Jennifer McKinney.

2.3.4. YAC fragility assay. As previously described (261,262) single colonies containing YACs with either the Z-DNA-forming or control B-DNA-forming sequence were used to inoculate 2 mL cultures in SD base with CSM-Leu media, and were grown for 20 hours at permissive temperature of 27-30°C, as several strains are temperature-sensitive and must be grown at lower temperatures. For each culture, 50 µL were plated on SD base with CSM-Leu plates containing 5-fluoroorotic acid (5-FOA) (Zymo Research, Irvine, CA) to select for mutagenic events resulting in the loss of functional *URA3* on the YAC. For each culture, 10 µL was diluted 1:10,000 and plated on SD base with CSM-Leu plates to serve as a total cell number count. The mutation frequencies were calculated as the number of FOA-resistant (FOA^R) colonies divided by the number of total colonies (Fig. 2.6). This assay was performed in triplicate on five separate clones for the Z-DNA-containing and control YACs for all wild-type and mutant yeast strains analyzed in this study. Student's *t*-test was used to determine significant differences between strains.

2.3.5. YAC mutation spectra analysis. For each strain containing the Z-DNA-containing or control B-DNA-containing YAC studied in the fragility assay, 30 FOA^R colonies were streaked onto SD base with CSM-leu+FOA plates to confirm FOA^R (Fig. 2.12). Single colonies were inoculated in water and cells were lysed by boiling at 95°C for 5 min, and used as template for PCR with the sets of primers listed in Table 2.1. The M13 primer set was specific to the Z-DNA-

forming or control B-DNA-forming region, the URA set serving as a measure of FOA^R, and the LEU set served as a loading control. Amplified DNA was separated on a 1% agarose gel, stained with ethidium bromide, and visualized on a ChemiDoc Imaging System (Bio-Rad, Hercules, CA).

2.3.6. Determination of mutation index (MI). We define the mutation frequency index (MFI) computed as the ratio between the mutation frequencies determined for the cells containing the Z-DNA-forming YAC, and that of cells containing the control B-DNA-forming YAC for each mutant yeast strain (Fig. 2.9d). MFI values from different experiments were pooled to obtain means and standard deviations. MSI, or mutation spectra index, was computed as follows. For each yeast strain we assessed the mutation spectra observed in 29-30 FOA^R colonies by calculating the percent of FOA^R colonies for each of the 4 types of mutations observed (described in Fig. 2.13b; Table 2.3). The percent values obtained with the strains harboring the control B-DNA-forming YAC were then subtracted from the values obtained with the corresponding strain harboring the Z-DNA-forming YAC. MSI corresponds to the sum of the absolute differences for each mutation category for each strain. The mutation index (MI) represents the product of MFI multiplied by MSI; specifically MI = MFI * (MSI + 1) / 10. Thus, when comparing among different yeast strains, the larger the MI values, the more pronounced are the differences in mutation frequencies and mutation spectra for the Z-DNA-containing clones relative to the control B-DNA-containing clones.

2.3.7. Chromatin immunoprecipitation (ChIP) assay in yeast. ChIP assays were performed using a Simple ChIP Enzymatic Chromatin IP Kit (Cell Signaling, Inc., Santa Cruz, CA) with some modifications to the manufacturer's protocol. Cultures of SD base CSM-Leu media were inoculated with a single colony of yeast cells containing YACs with the Z-DNA-forming or control sequence, and were grown overnight at permissive temperatures. Cultures were fixed with 1% formaldehyde for 10 min at room temperature to crosslink DNA and proteins. Crosslinking was quenched by adding 10X Glycine and incubating for 5 min at room temperature. Cells were washed three times with cold PBS (Corning Inc., Corning, NY) containing 1 mM cold phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich, St. Louis, MO) (100 mM in 100% ethanol). The yeast cells were lysed in buffer [0.1 M Tris-Cl pH 8.0, 50 mM ethylenediaminetetraacetic acid (EDTA), 1% sodium dodecyl sulfate (SDS)] and Zymolyase (Zymo Research) with rotation at 37°C for 1 hour. To ensure cell lysis, sterile, acid washed glass beads (RPI Corp., Mount Prospect, IL) were added to the cells and the samples were vortexed for a total of 5 min, alternating 30 sec on vortex, 30 sec on ice. The supernatant was transferred to a new tube, and sonicated to obtain an average DNA fragment size of <800 bp. Fragment size was confirmed following the ChIP kit protocol, and fragments were separated on a 1% agarose gel, stained with ethidium bromide, and visualized on a ChemiDoc Imaging System. The fragmented chromatin was aliquoted for incubation with primary antibody, and 2% of the chromatin was stored as input DNA. Aliquots were incubated with

5 µg of antibody (α-IgG antibody (ChIP kit) and α-H3 antibody, (Abcam 8895)) at 4°C with rotation overnight. The remainder of the assay was performed following the ChIP kit protocol. Purification of the DNA was performed using a PCR Purification Kit (Qiagen). Fractions of purified ChIP and input DNA were used for PCR analysis. Primer sets for PCR amplification included M13-20 and M13Rev for specificity surrounding the Z-DNA-forming or control B-DNA sequence and Leu2S2For and Leu2S2Rev as a non-specific control within the *LEU2* gene (Table 2.1). Amplified products were separated on a 1% agarose gel, stained with ethidium bromide, and visualized on a ChemiDoc Imaging System. ImageJ software was used to quantify band intensity.

2.3.8. Z-DNA-induced mutagenesis assay in human cells. Isogenic human XPF-proficient (GM08437B-XPF) or XPF-deficient cells (GM08437B-pLPC) (266,267) were maintained in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, CA) with 10% fetal bovine serum (FBS) and antibiotics. The Z-DNA-containing (pU$(CG)_{14}$) (CGCGCGCGCGCGCGCGCGCGCGCGCG) or control B-DNA-containing plasmid (pUCon) (CGAGCTATCTGAGTCGAATACAGTTCGAC) was transfected into human cells using GenePORTER (Genlantis Inc., San Diego, CA) according to the manufacturer's recommendations using 3 µg of plasmid DNA. Cells were harvested at 48 hours following transfection, and plasmid DNA was extracted using Hirt's method (268) with slight modifications as previously described (262). Purified plasmid DNA was

treated with DpnI digestion as previously described to remove plasmids that were not replicated in the mammalian cells (262), and again purified using phenol-chloroform extraction and ethanol precipitation prior to transformation via electroporation into *E. coli* DH10β electrocompetent bacterial cells (New England Biolabs, Inc., Ipswich, MA), which were used in the *lacZ*-based blue/white screening assay to determine mutation frequencies as previously described (262). Experiments were performed in triplicate and Student's *t*-test was used to calculate statistical values. DNA from individual mutant colonies was sequenced to characterize the mutations/deletions near the Z-DNA or control B-DNA regions.

2.3.9. siRNA knockdown of MSH2 in human cells. Human MSH2 siRNA or non-targeting siRNA (Dharmacon, GE Healthcare) (Table 2.1) was transfected into cultured human XPF-proficient cells (GM08437B-XPF) with GenePORTER (Genlantis Inc., San Diego, CA) using the manufacturer's recommended protocol at a final siRNA concentration of 0.1 μM. A second siRNA transfection together with the pU$(CG)_{14}$ and pUCon mutation reporter plasmids was carried out 48 hours (T_{48}) after the first transfection using GenePORTER according to manufacturer's recommended protocol at a final siRNA concentration of 0.1 μM, and 3 μg of plasmid DNA. Cells were harvested at T_{48} as well as 16 or 48 hours after the second transfection (T_{64} and T_{96}) as indicated for Western blotting to

confirm the knockdown of MSH2, and plasmids were extracted for analysis by mutagenesis and ChIP assays.

2.3.10. Determination of siRNA knockdown by Western blotting. XPF-proficient cells (GM08437B-XPF) were harvested 48 hours after plasmid transfection. Cells were lysed in 1x radioimmunoprecipitation assay (RIPA) buffer (with proteinase inhibitor cocktail) on ice for 1 hour, followed by sonication. A portion of the supernatant (equivalent to 10-30 µg of protein) was mixed with 2x SDS loading buffer and boiled for 5 min at 95°C. Samples were separated by SDS-PAGE on a 4-15% Criterion™ TGX™ (Tris-Glycine-eXtended) midi gel, and transferred to a polvinylidene difluoride (PVDF) membrane using a Trans-Blot® Turbo™ Transfer System (Bio-Rad, Hercules, CA). The membrane was blocked for 1 hour at room temperature with 1x TBST (tris buffered saline (0.02 M tris base, 0.15 M NaCl) plus 0.05% Tween 20) with 5% blotting grade dry milk (Bio-Rad), and probed with antibodies against MSH2 (1:100, Calbiochem, Billerica, MA), XPF (1:1000, Cell Signaling), and PCNA (1:10,000, Santa Cruz), and visualized on a ChemiDoc Imaging System. ImageJ software was used to quantify band intensity.

2.3.11. Chromatin immunoprecipitation (ChIP) assay in human cells. Detection of protein enrichment at a Z-DNA-forming site was performed as previously described with some modifications (130,269). Briefly, human cells

were plated in antibiotic-free Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, CA) with 10% fetal bovine serum (FBS) at a concentration of 240,000 cells/plate in 60 mm plates. The next day, 3 μg of pU$(CG)_{14}$ or pUCon mutation reporter plasmids were transfected into the cells using GenePORTER (Genlantis Inc., San Diego, CA) according to manufacturer's suggested protocol. At 16 hours post-transfection, the cells were fixed with formaldehyde at a final concentration of 1% for 10 min at room temperature. The remainder of the ChIP assay was performed using Simple ChIP Enzymatic Chromatin IP Kit (Cell Signaling, Inc., Santa Cruz, CA), following manufacturer's protocol with some modifications. Aliquots of the chromatin were incubated with 1 μg of primary antibody (α-IgG and α-H3 antibodies (ChIP kit), α-MSH3 antibody (Thermo Fisher), α-MSH2 antibody (Calbiochem), α-MSH3 antibody (Thermo Scientific), α-MSH6 antibody (Abcam), α-XPA antibody (Abcam), and α-XPF antibody (generous gift from Dr. Richard Wood, University of Texas M.D. Anderson Cancer Center) at 4°C with rotation overnight. The remainder of the assay was performed following the ChIP kit protocol. DNA fragments were purified using Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, WI). Fractions of purified ChIP and input DNA were used for PCR analysis. Primers used for PCR amplification included pUinsFor1 and pUinsRev1 (Table 2.1). Amplified products were separated on a 1% agarose gel, stained with ethidium bromide, and visualized on a ChemiDoc Imaging System. ImageJ software was used to quantify band intensity. Experiments were

performed in triplicate and Student's *t*-test was used to calculate statistical values.

2.3.12. Confirmation of Z-DNA structure formation in plasmids by S1 nuclease sensitivity assays. As previously described (270) with slight variations, Z-DNA-containing or B-DNA-containing control plasmids (2 µg) (pUCG14 and pUCon, respectively) were incubated at room temperature for 5 min with 50 mM NaCl and 4 mM $MgCl_2$, then digested with 2 µL S1 nuclease (Promega Corporation, Madison, WI) in 25 µL reactions in provided S1 reaction buffer for 20 min at 37°C. Reactions were stopped by heat inactivation at 65°C for 20 min and diluted to 100 µL. Plasmid DNA was extracted using phenol chloroform, and precipitated using ethanol. For some experiments, purified plasmid DNA was further digested with a combination of 1 µL each of XmnI, and/or AflII endonucleases (at position 93, and 1716 respectively; Fig. 2.20) in 20 µL reactions for 3 hours at 37°C. Reactions were stopped by heat inactivation at 65°C for 20 min. Cleaved products were separated on 1.4% agarose gels, stained with SYBR® Gold (Life Technologies), and visualized on a ChemiDoc Imaging System. ImageJ software was used to quantify band intensity. Experiments were performed in triplicate.

2.3.13. *In vitro* DNA cleavage assay via primer extension. Isogenic human XPF-proficient (GM08437B-XPF) or XPF-deficient (GM08437B-pLPC) whole cell

extracts (WCE) were prepared from human XPF$^{+/-}$ cells (described in section 2.3.8) using Nucbuster™ Protein Extraction Kit (EMD Millipore, Temecula, CA) starting with Reagent 2. Reaction mixtures contained 600 ng of Z-DNA-forming or control B-DNA-forming plasmid DNA (pU(CG)$_{14}$ and pUCon, respectively), and were incubated at room temperature for 5 min with 50 mM NaCl and 4 mM $MgCl_2$. Samples were then incubated in reaction buffer (5 mM $MgCl_2$, 40 mM Hepes-KOH pH 7.8, 0.5 mM DTT, 2 mM ATP, 22 mM phosphocreatine, 0.36 mg/mL BSA, 50 ng/µL CPK, 30 mM KCl), and water or 100 ng of WCE for 30 min at 30°C. After incubation, samples were treated with 0.02 M ethylenediaminetetraacetic acid (EDTA) and 80 µg/mL RNAse A and further incubated at 37°C for 10 min. Samples were then treated with 0.5 % SDS and 1 mg/mL proteinase K and incubated at 65°C for 30 min. Plasmid DNA was purified using phenol-chloroform and ethanol precipitation, resuspended in nuclease-free water and used as template for PCR with JMLeft or JMRight primer (Table 2.1 and Fig. 2.21) to detect nicks on the plasmid DNA. Amplified DNA was separated on a 1.5% agarose gel, stained with SYBR® Gold (Thermo Fisher Scientific, Waltham, MA) and visualized on a ChemiDoc Imaging System (Bio-Rad, Hercules, CA). Experiments were performed in triplicate.

2.3.14. Analysis of Z-DNA-induced DSBs by linker-mediated PCR (LM-PCR). LM-PCR analysis was performed as previously described by our laboratory (125,127,262). Briefly, 8 µg of pUCon (control B-DNA) or pU(CG)$_{14}$ (Z-

DNA) plasmid DNA was transfected into isogenic human XPF-proficient (GM08437B-XPF) or XPF-deficient (GM08437B-pLPC) cells using GenePORTER (Genlantis Inc., San Diego, CA) according to manufacturer's suggested protocol. At 24 hours post-transfection, plasmids were recovered and purified using Hirt's method (268) with slight modifications as previously described (262). Purified plasmid DNA was treated with Pol I Klenow fragment (New England Biolabs, Inc., Ipswich, MA) according to manufacture's recommendations to create blunt-ended DNA, and plasmid DNA was purified by phenol-chloroform extraction and ethanol precipitation. Purified Klenow-treated DNA was ligated to linker DNA containing one blunt end and one 3' overhang (product of annealed 5' phosphorylated LMPCR1 and LMPCR2 primers, Table 2.1) at 16°C overnight, followed by 20 min at 70°C, and allowed to cool to room temperature. Ligation products were used as templates in the PCR reactions to amplify the region between the linkers and the upstream primer 190 bp upstream of the $(CG)_{14}$ insert (LacZleft741 and LMPCR2, respectively, Table 2.1). PCR products were separated on a 1.2% agarose gel, stained with ethidium bromide and visualized on a ChemiDoc Imaging System (Bio-Rad, Hercules, CA).

2.4. Results

2.4.1. Z-DNA-forming sequences map to breakpoints in human cancer genomes.

To determine the impact of Z-DNA-forming sequences in human cancer etiology, our laboratory recently searched for Z-DNA-forming repeats (ZFRs) within ±100 bp (thereafter referred as to bins) of 19,956 human cancer genome translocation breakpoints (database from COSMIC at http://cancer.sanger.ac.uk/). We determined the distributions of each ZFR midpoint relative to the breakpoint positions (taken to be 0). ZFRs were significantly enriched in regions within ±100 bp of the 19,967 cancer translocation breakpoints, compared to the 20,000 randomly picked loci that served as the control, with an obvious peak centering at the breakpoint in the cancer genomes (166) (Fig. 2.3). Thus, Z-DNA-forming sequences are significantly enriched surrounding human cancer translocation breakpoints, implicating Z-DNA in chromosomal instability and cancer development.

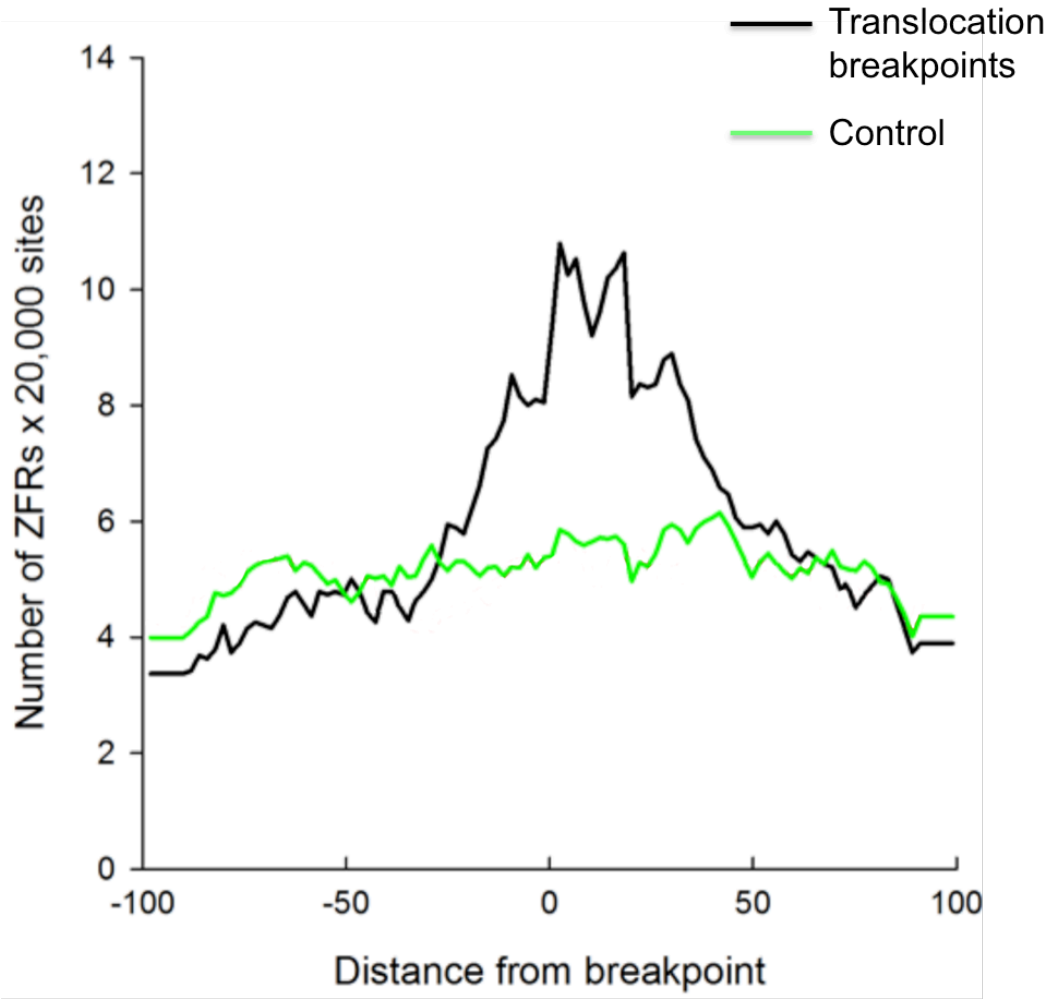


Figure 2.3. Enrichment of Z-DNA forming repeats (ZFRs) at human cancer translocation breakpoints. Normalized numbers of ZFRs found at cancer translocation breakpoints (black) and random loci (green) are shown. These data were obtained by our laboratory [variation of data from Bacolla, *et. al. Nucleic Acids Res* 2016 (166)].

2.4.2. Z-DNA-induced chromosomal instability in yeast.

We, and others, have shown that Z-DNA-forming sequences can cause genomic instability in bacteria, human cells, and mice, and can alter DNA metabolism (i.e., DNA transcription, replication, and repair) (11,127,128,257). To determine if a Z-DNA-forming sequence was mutagenic in yeast, we transformed a yeast artificial chromosome (YAC) containing a Z-DNA-forming $(CG)_{14}$ sequence or a control B-DNA-forming sequence (control) (Fig. 2.4) into wild-type BY4742 cells via *kar*-mediated transfer (Fig. 2.5, Table 2.2), and screened for 5-Fluoroorotic acid-resistant (FOA^R) colonies as a measure of mutation frequency (Fig. 2.6). We demonstrated for the first time that the $(CG)_{14}$ Z-DNA-forming sequence stimulated significantly higher mutation frequencies in wild-type yeast when compared to the control B-DNA-forming sequence (18.6-fold over control; Fig. 2.7).

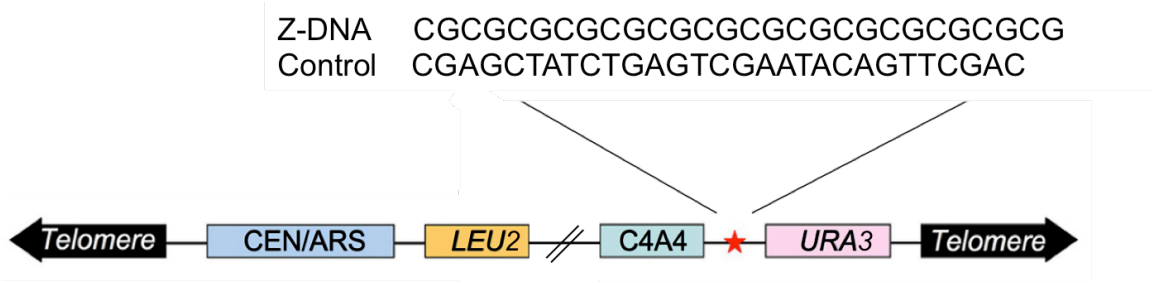


Figure 2.4. Schematic diagram of the Yeast Artificial Chromosome (YAC) used in this study. Representation of the YAC used in these studies (not to scale). The full-length YAC is ~62 kb in length, containing a yeast origin of replication (CEN/ARS), a telomere seed (C4A4), and two selective genes (*LEU2* and *URA3*). Thus, cells grown in CSM-leu-ura selective media will only grow if they contain the YAC. The red star represents the location of either the control B-DNA-forming or Z-DNA-forming sequence (shown above the YAC). There is ~41 kb of lambda DNA between the *LEU2* gene and the telomere seed [adapted from (262)].

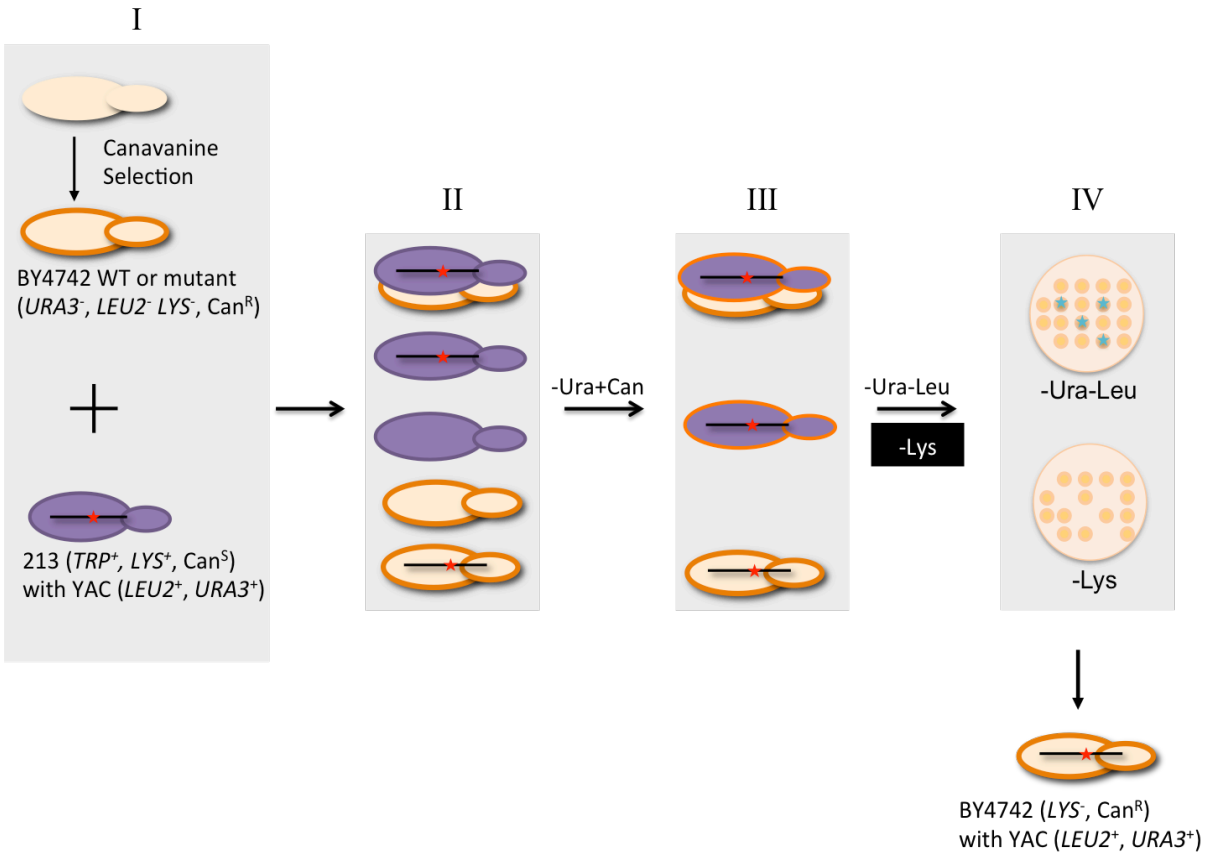


Figure 2.5. *Kar*-mediated transfer of YAC from donor to recipient cells. The 213 donor cell line (purple) containing the YAC is *Can*^{S}, while the recipient BY4742 WT or mutant cells (peach) are *Can*^{R}, which is verified prior to the *kar* cross by plating on YPD plates with canavanine. Liquid cultures containing either donor or recipient cells are mixed and inoculated into YPD media and incubated for 6 hours at permissive temperature **(I)**, resulting in various progeny **(II)**. Cells are plated onto CSM-ura+canavanine to select for only those cells with the BY4742 background that contain the YAC **(III)**. To ensure cells are haploids that only contain the BY4742 background and are not a result of mating between recipient and donor, *Can*^{R} colonies are replica-plated onto both CSM-leu-ura and CSM-lys plates **(IV)**. Colonies that grow on the CSM-leu-ura plates, but do not grow on CSM-lys plates (blue stars), are selected as BY4742 background containing the YAC (*LYS*$^{-}$, *Can*^{R}, *LEU2*$^{+}$, *URA3*$^{+}$). The black line represents the YAC, and the red star represents the control or Z-DNA sequence (Fig. 2.4).

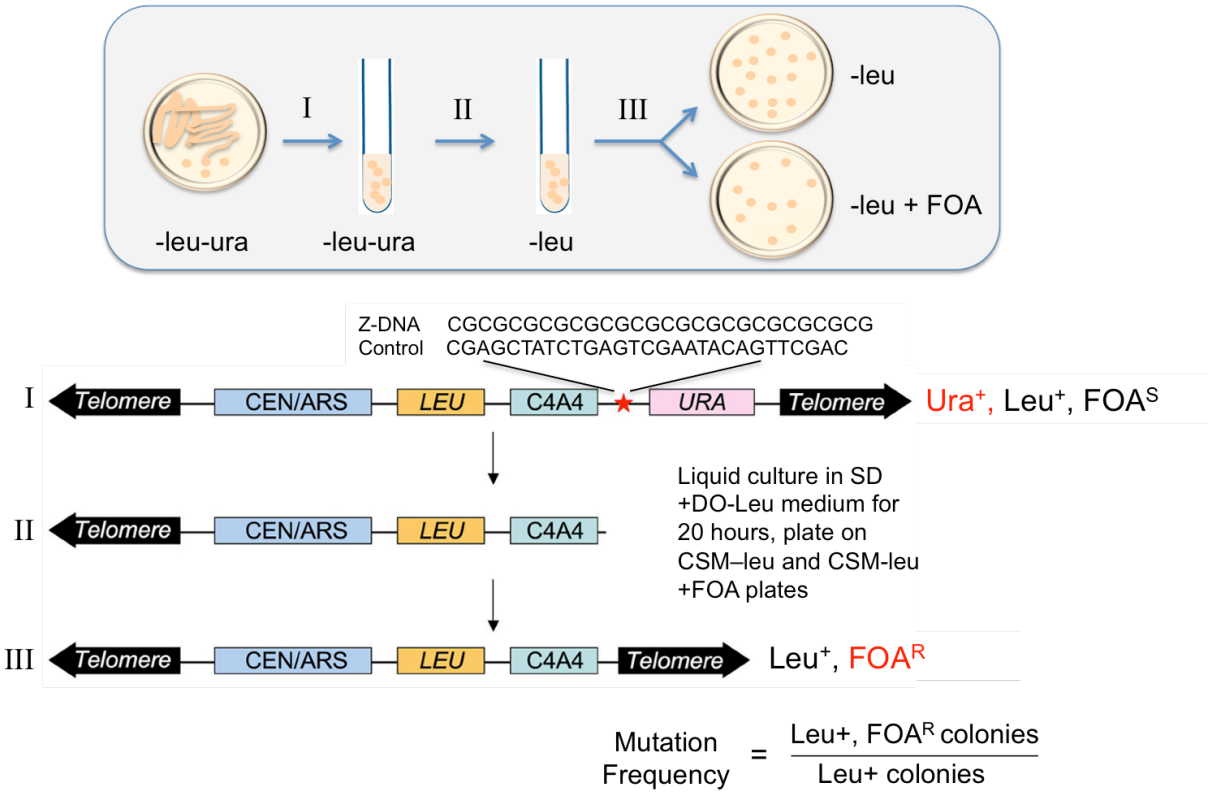


Figure 2.6. YAC fragility assay. A single colony containing the full-length, *URA3*$^{+}$ YAC is harvested from a CSM-leu-ura plate, inoculated into 2 mL CSM-leu-ura liquid culture, and grown for 24 hours at permissive temperature **(I)**. Cells from the CSM-leu-ura culture are inoculated into 2 mL CSM-leu cultures, and grown for 20 hours at permissive temperature **(II)**. By allowing the cells to grow in CSM-leu only selective media, mutations can occur in the *URA3* gene, resulting in a mutated YAC that will become FOA^{R}. CSM-leu media selects for cells containing either the full-length wild-type YAC, or a mutated YAC. Cells are then plated onto CSM-leu+FOA plates to select for cells containing a mutated YAC, as well as diluted 10,000-fold and plated onto CSM-leu plates as a measure of total cell number **(III)**. Mutation frequencies are calculated as a ratio of the number of colonies on the FOA plates (mutants) to the number of colonies on the –leu plates (total cell number) [adapted from (261-263)].

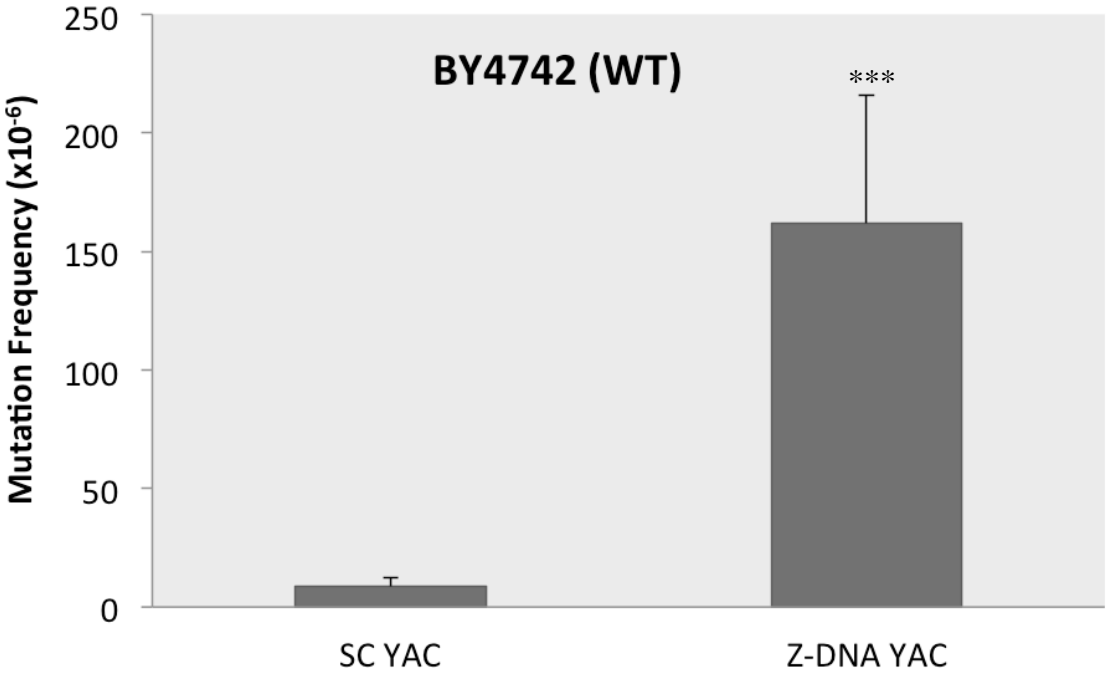


Figure 2.7. Z-DNA is mutagenic in wild-type yeast. YAC fragility assay performed on BY4742 wild-type (WT) cells containing either a control B-DNA-containing YAC (SC) or a YAC Z-DNA-containing YAC (Z-DNA). Mutation frequencies were measured as number of FOA^R colonies divided by total cell number. [*P* value from two-sided Student's *t*-test $<1x10^{-5}$ (***)]. These experiments were performed by Jennifer McKinney.

To identify gene products involved in Z-DNA-induced mutagenesis in yeast, we obtained an *S. cerevisiae* genome library deficient in non-essential genes from American Type Culture Collection. The deletion library allowed us to screen multiple knockout strains for a variation of the effect on the induction of mutagenesis we see in wild-type cells caused by Z-DNA using the YAC fragility assay. By using this assay on mutant strains and comparing mutation frequencies of mutant cells with the Z-DNA-containing YAC versus the same mutant cells with the control B-DNA-containing YAC, we were able to identify proteins that are involved in Z-DNA-induced mutagenesis (Fig. 2.8). For example, if the Z-DNA-induced mutagenesis in a mutant strain is *increased* compared to the induction seen in WT cells, the protein that is deficient in the mutant strain may be involved in preventing Z-DNA formation or in error-free repair of the Z-DNA structure (Fig. 2.8a and b). Alternatively, if the induction caused by Z-DNA is *decreased* compared to WT cells, the protein may be involved in promoting Z-DNA formation, stabilizing the Z-DNA structure following its formation and subsequently blocking error-free repair, or recognizing or processing the Z-DNA structure resulting in mutations via error-prone repair (Fig. 2.8c-e). Finally, it is possible that the same induction seen in WT cells would be seen in a mutant strain, suggesting the protein that is deficient in the mutant strain may not play a role in Z-DNA-induced mutagenesis, and further studies would be needed to confirm this conclusion.

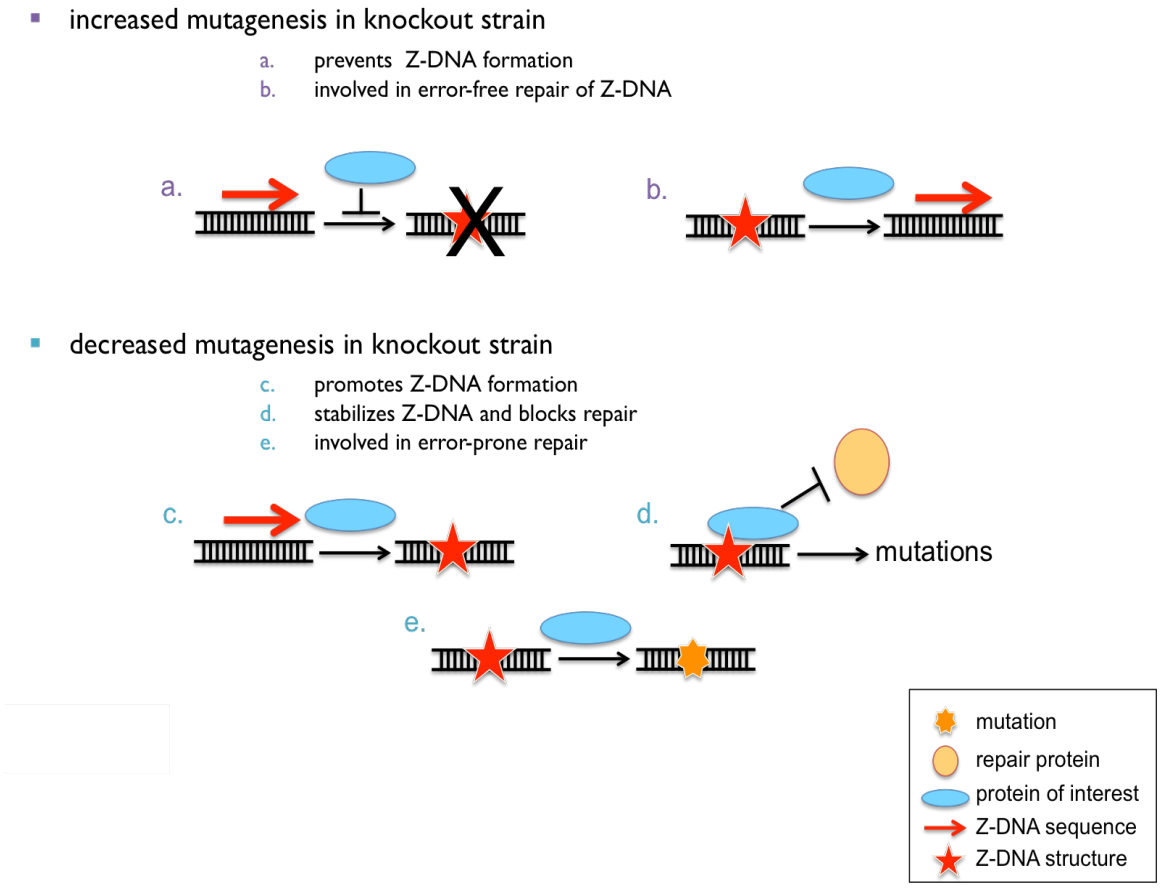


Figure 2.8. Potential outcomes of YAC fragility assay. This assay was used to measure Z-DNA-induced mutagenesis in wild-type cells and mutant cells. There are several possible outcomes of the fragility assay shown here. **(a-b)** If there is an increased mutation frequency in the Z-DNA-containing YAC (compared to wild-type cells) in the deficient cells, then this may indicate that the protein either prevents the non-B structure from forming, or it repairs the structure in an error-free manner. **(c-e)** However, a decreased mutation frequency may indicate that the protein promotes the formation of non-B-DNA, or stabilizes the structure once formed and prevents repair by other proteins, or attempts to repair the structure in an error-prone manner leading to subsequent mutations.

The helical distortions induced by Z-DNA formation and single-stranded regions at the B-Z junctions mimic substrates for the NER and MMR pathways. Therefore, we predicted that Z-DNA might also be recognized and/or processed by components of such pathways; thus we screened NER and MMR proteins as well as other structure-specific endonucleases for an effect on Z-DNA-induced mutagenesis. Each mutant strain was transformed with the control B-DNA-containing (control) or Z-DNA-containing YAC via *kar*-mediated transfer (Fig. 2.5). We performed the YAC fragility assay (Fig. 2.6) on all mutants listed in Table 2.2.

When comparing absolute values, there were significantly more mutations in cells containing the YAC with the Z-DNA-forming sequence compared to the control B-DNA-forming YAC in most strains tested, similar to the effect seen in WT strains (Fig. 2.9). Interestingly, the most striking differences from wild-type cells occurred in the NER mutants *rad1Δ(XPF)* and *rad10Δ(ERCC1)* (Fig. 2.9a), MMR mutants *msh2Δ* and *msh3Δ* (Fig. 2.9b), and structure-specific nuclease mutants *mre11Δ and sae2Δ* (Fig. 2.9c; *P* value <0.001) (e.g., a decrease of mutation frequency to background levels), suggesting that in the absence of these proteins, Z-DNA is no longer mutagenic in this assay.

Due to the inherent increased level of spontaneous background mutations in several of the repair mutant strains, we also considered the fold induction of the Z-DNA-containing YAC compared to the control B-DNA-containing YAC within each strain. Compared to the 18.6-fold induction seen in WT, there was an

overall significant decrease in the fold induction seen within each mutant set comparing the Z-DNA YAC to control B-DNA YAC (Fig. 2.9d and Table 2.3). Again, the most dramatic loss of effect caused by Z-DNA in the WT strains occurred in the *rad1Δ(XPF), rad10Δ(ERCC1), msh2Δ, msh3Δ, mre11Δ, and sae2Δ* mutant strains. In contrast to the overall *decrease* seen in the fold induction within the mutant strains compared to the induction seen in WT (Fig 2.9d), subtracting the background from the Z-DNA-induced mutation frequency demonstrated a significant *increase* in Z-DNA-induced mutagenesis in the *rad14Δ(XPA), rad2Δ(XPG), rad27Δ(FEN1), and mlh1Δ* strains (Fig. 2.10 and Table 2.3). In accordance with the absolute and fold data, *rad1Δ(XPF), rad10Δ(ERCC1), msh2Δ, msh3Δ* mutants display a highly significant decrease in Z-DNA-induced mutagenesis compared to that of WT, that was not detected in other mutant strains within the respective pathways. Regardless, whether we considered absolute values, fold induction, or values with background subtracted, it was clear that the dramatic effect seen in these mutants was not observed in other mutants within the same pathways; thus, these data suggest that proteins from multiple pathways are involved in Z-DNA-induced mutagenesis.

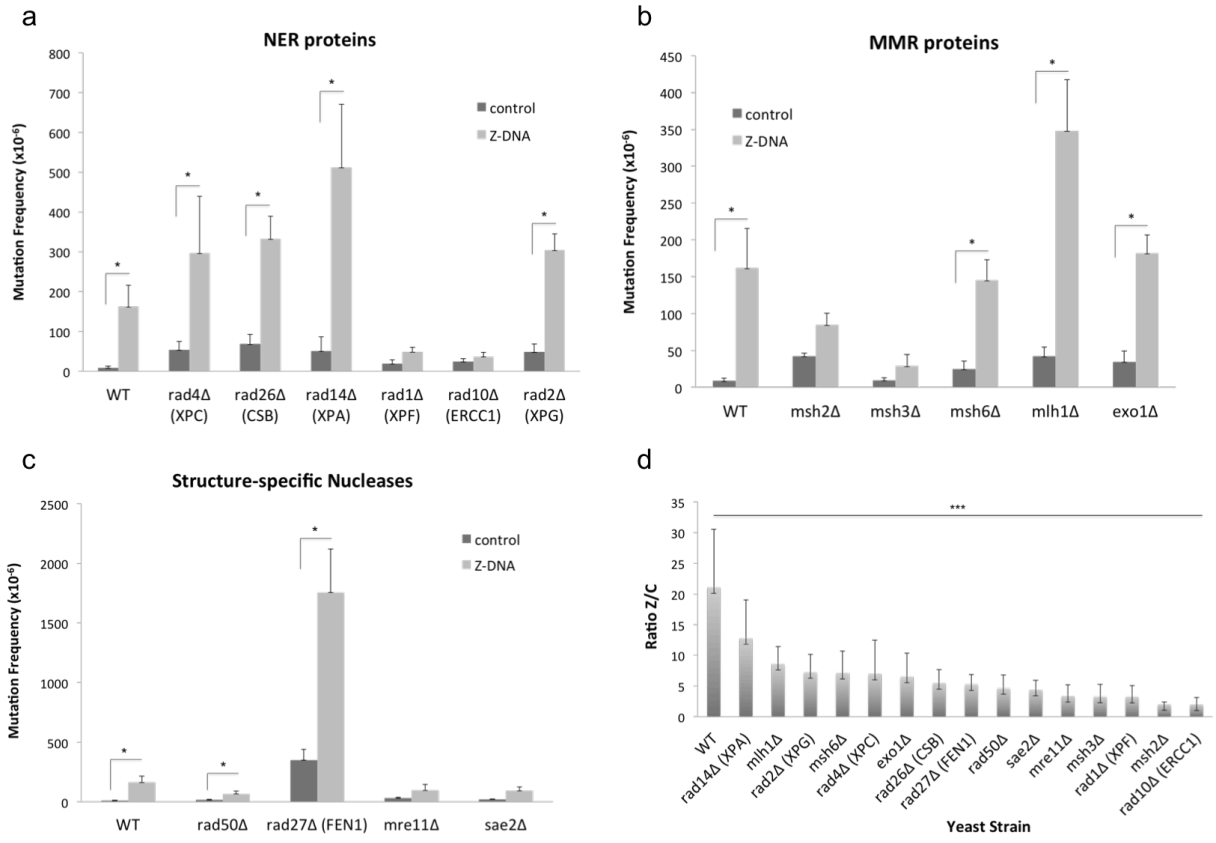


Figure 2.9. Z-DNA-induced mutagenesis in wild-type and repair-deficient yeast cells. The Z-DNA-forming $(CG)_{14}$ insert resulted in increased mutation frequencies compared to the control B-DNA-forming insert in BY4742 wild-type yeast cells deficient in NER **(a)**, MMR **(b)**, and other repair proteins including structure-specific nucleases **(c)**. However, the fold increase of the $(CG)_{14}$ insert compared to control insert within each mutant set was decreased compared to the fold increase in BY4742 wild-type cells **(d)**. [Student's *t*-test was used to calculate *P* value <0.001 (*), $<1x10^{-10}$ (***)]. These experiments were performed by Jennifer McKinney. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

Table 2.3. Mutation frequencies listed as a ratio (Z/C) and with background subtracted (Z-C).

	Yeast Strain	Z/C	Z-C ($x10^{-6}$)
WT	BY4742	18.6	153.2
NER	*rad4Δ(XPC)*	5.6	243.1
	rad26Δ(CSB)	4.9	263.8
	rad14Δ(XPA)	10.1	461.2
	rad1Δ(XPF)	2.6	29.6**
	rad2Δ(XPG)	6.3	255.7
	rad10Δ(ERCC1)	1.7	15.1**
MMR	*msh2Δ*	2.0	42.5**
	msh3Δ	3.0	19.3**
	msh6Δ	5.9	120.7
	mlh1Δ	8.2	305.9
	exo1Δ	5.2	147.2
Structure-specific endonucleases	*rad27Δ(FEN1)*	5.0	1404.6
	mre11Δ	3.1	65.1
	sae2Δ	4.3	71.0
	rad50Δ	4.2	50.9**

P value $<1x10^{-4}$ compared to WT (**)
[McKinney, *et al.* (submitted *Molecular Cell* 2016)]

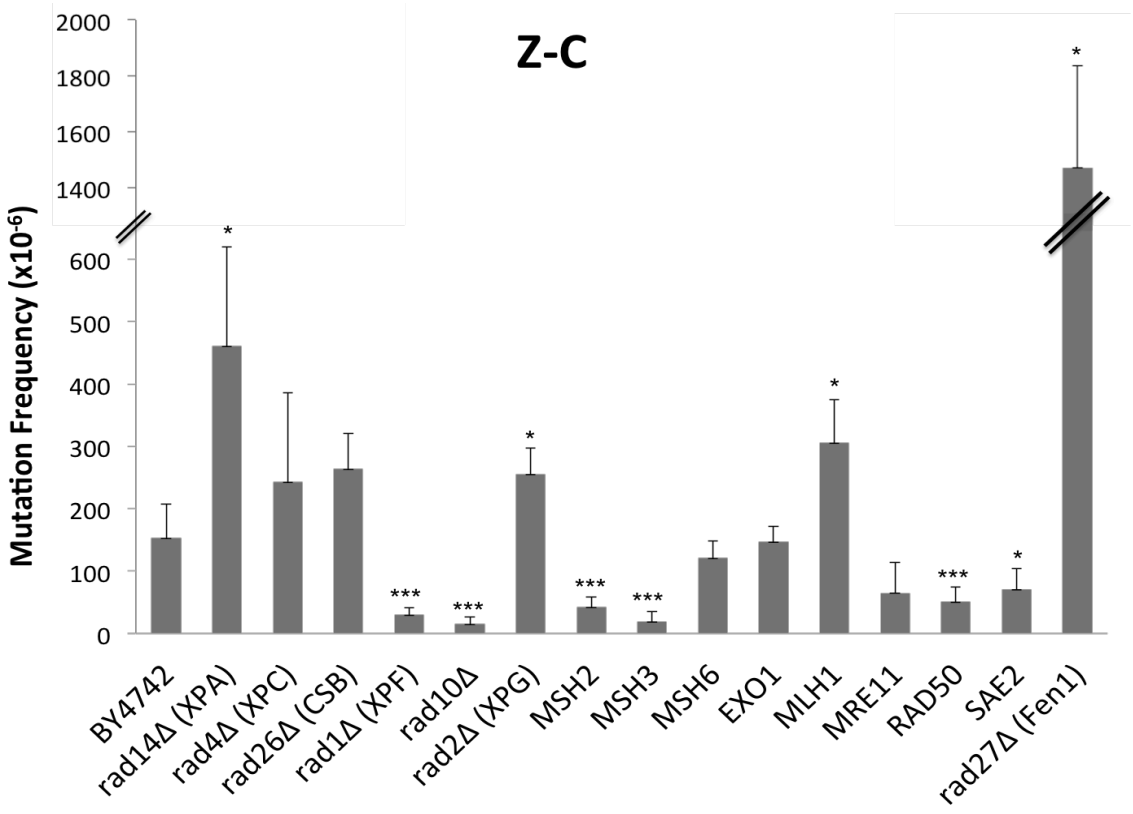


Figure 2.10. Z-DNA-induced mutagenesis with background subtracted. Raw data from Fig. 2.9a-c with background mutations (i.e. spontaneous mutations) subtracted results in the observation of several strains with a significant difference from that of WT [*P* value <0.01 (*), $<1x10^{-4}$ (***)].

2.4.3. Mutation spectra of spontaneous and Z-DNA-induced mutations in yeast.

The *URA3* gene encodes orotidine-5'phosphate decarboxylase (ODCase), an enzyme required for the biosynthesis of pyrimidine ribonucleotides. In strains expressing the functional *URA3* gene, ODCase converts 5-FOA to the toxic form of 5-flurouracil (271-273) (Fig. 2.11). Thus, the YAC fragility assay evaluates FOA^R as a measure of the frequency of mutations that occur due to the loss of the *URA3* gene on the YAC. We are most interested in frank DSBs that occur near Z-DNA structure formation, resulting in the loss of the right arm of the YAC, as these events may relate to translocations. However, there are other events that can result in loss of a functional *URA3* gene and cause the cells to become FOA^R. Any mutation, including point mutations and small or large deletions within the *URA3* gene result in FOA^R.

To determine if FOA^R is due to complete loss of the right arm of the YAC, caused by a DSB, or due to inactivation of the *URA3* gene by a point mutation, we analyzed FOA^R colonies by PCR using primers that amplify the Z-DNA or control B-DNA sequence, a region of the *URA3* gene, and a region of the *LEU2* gene (Fig. 2.12a and b; Table 2.1). All three primer sets were verified to be specific to the YAC, and did not amplify any region in the yeast wild-type or mutant genome (Fig. 2.12c).

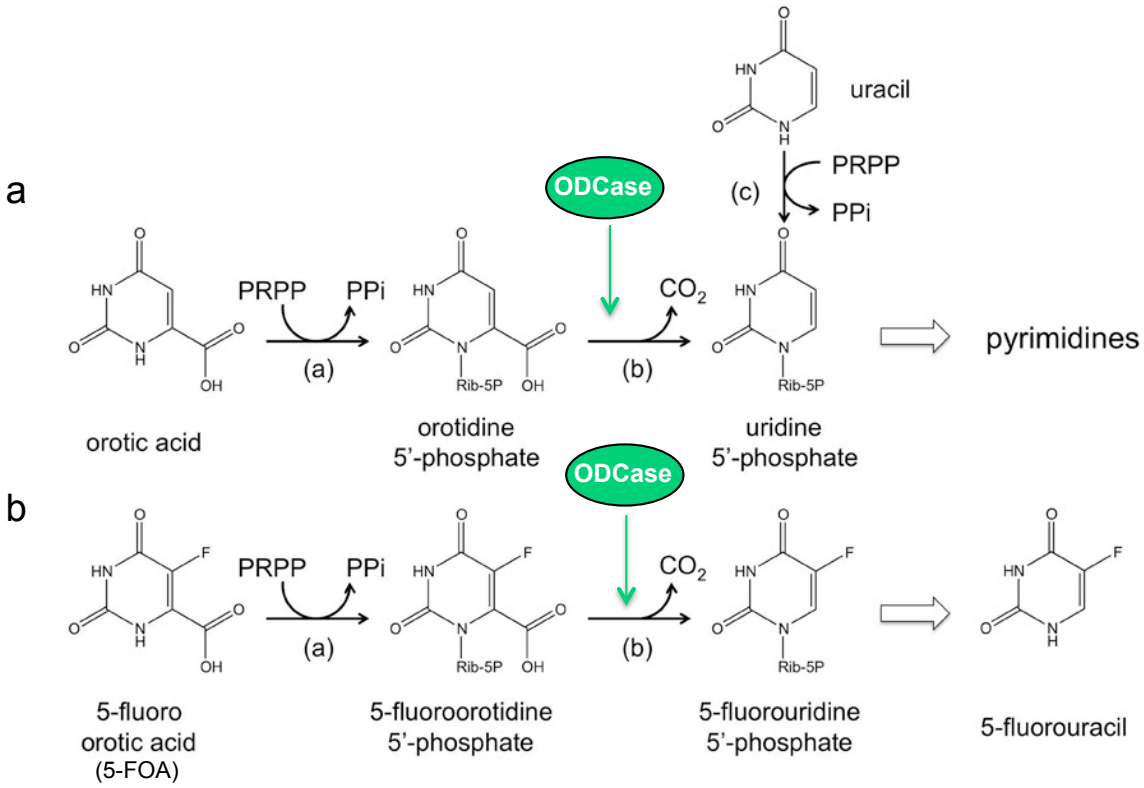


Figure 2.11. 5-FOA is toxic to cells expressing *URA3* gene. (a) The *URA3* gene encodes for the orotine-5-monophosphate decarboxylase (ODCase) enzyme that is required for conversion of orotidine 5'phosphate to uridine 5'phosphate in the biosynthesis of pyrimidine ribonucleotides. **(b)** Functional ODCase will also result in the conversion of 5-fluorotic acid (5-FOA) to the toxic 5-flurouracil, resulting in cell death. Thus, cells that have lost a functional *URA3* gene become resistant to 5-FOA, as 5-FOA itself is nontoxic to the cells [adapted from (273)].

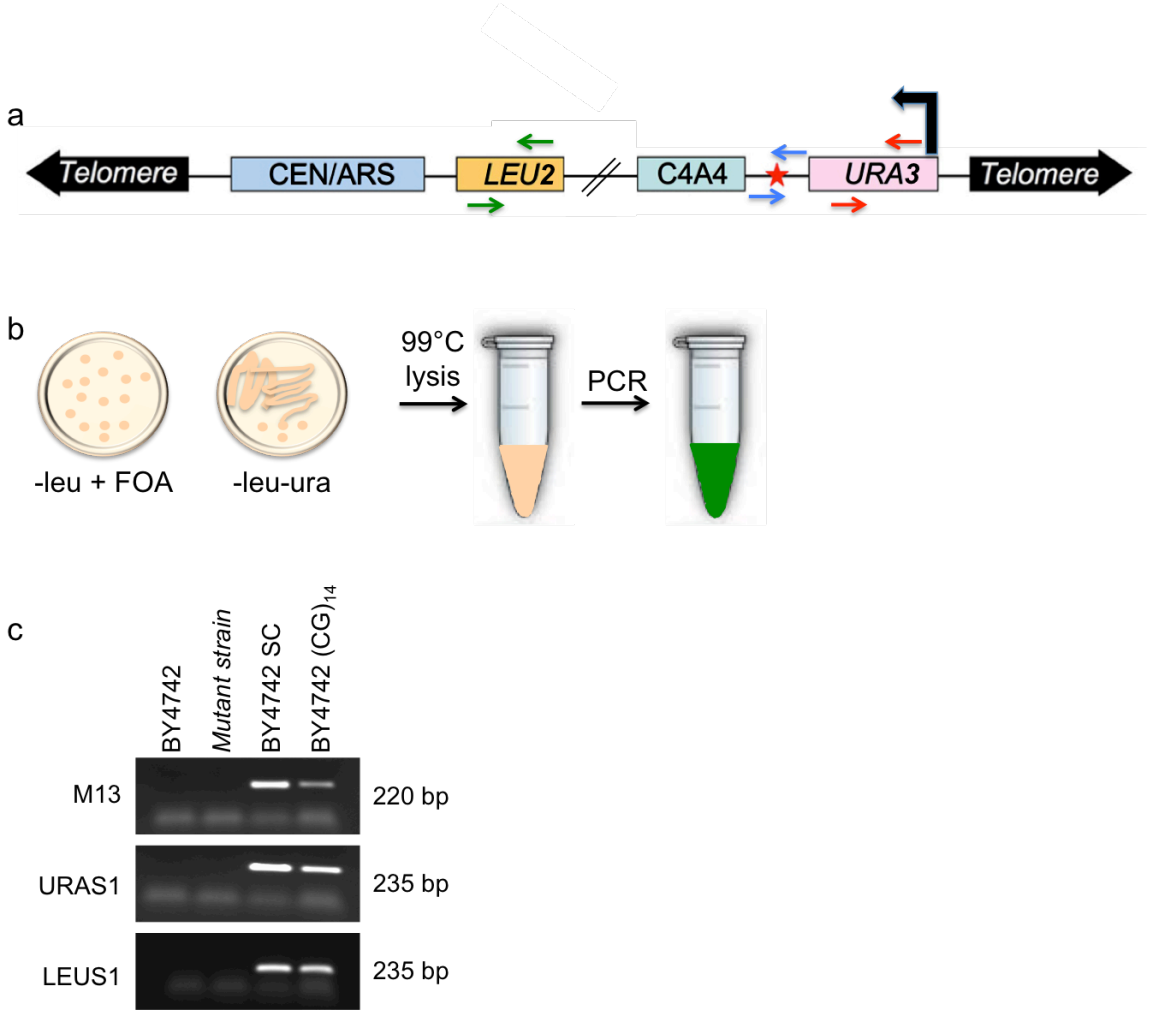


Figure 2.12. Primers used in this study are specific to the YAC. (a) To determine the type of mutation that occurred in the *URA*3 gene to result in FOA^R, we designed several primers that are specific to the *URA3* gene (red arrows), Z-DNA or control B-DNA insert (blue arrows), or the distal region in the *LEU2* gene, which also served as a loading control (green arrows). The black arrow represents the direction of the *URA3* promoter. **(b)** For the mutation spectra, single colonies from the FOA-containing or CSM-leu-ura plates were harvested, lysed at 99°C for 5 min, and used as template for PCR reaction. **(c)** We tested this assay using BY4742 and a mutant strain that did not contain the YAC to verify that the primers did not amplify any region in the yeast genome, and were specific to the YAC, as well as the BY4742 wild-type strains containing the control or $(CG)_{14}$ Z-DNA-forming YAC. The lighter bands on the gel are primer dimers. These experiments were performed by Jennifer McKinney [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

We analyzed 30 FOA^R colonies for each mutant and wild-type strain using all three primer sets as described above. Results obtained from PCR and gel electrophoresis were categorized according to potential causes of FOA^R as follows: 1) samples with amplification from all three sets of primers were categorized as having a point mutation or small deletion within the *URA3* gene; 2) samples with amplification from the M13 and LEU primer sets were categorized as having a deletion within the *URA3* gene; 3) samples with amplification from the URA and LEU primer sets were categorized as having a point mutation or small deletion in the *URA3* gene; and 4) samples with amplification from only the LEU primer set were categorized as having a double-strand break (DSB) and loss of the right arm of the YAC.

In the wild-type strain, there was a clear distinction between the control B-DNA-containing YAC, in which the majority of the mutants were point mutations/small deletions, and the Z-DNA-forming YAC, in which the majority of the mutants resulted from frank DSBs and loss of the right arm of the YAC (Fig. 2.13a). There were several mutant strains that demonstrated a mutation spectra similar to WT, in which there was a substantial increase in the number of DSBs and arm loss in the Z-DNA-forming YAC compared to the control B-DNA YAC, including *Δmsh6* and *Δrad26(CSB)* strains (Fig. 2.13c and d, respectively). In contrast, the mutation spectra for several mutants strains deviated from that of WT. For example, in *Δrad10(ERCC1)*, *Δmsh2,* and *Δmsh3* strains, the mutation spectra of the Z-DNA-forming YAC compared to the control B-DNA-forming YAC

were quite similar, suggesting that the mutations that occurred in cells lacking these proteins, were not influenced by Z-DNA-formation (Fig. 2.13c and d).

Interestingly, the mutation spectra of the *Δrad1(XPF)* strain demonstrated a difference between the control B-DNA and Z-DNA YACs. While the control YAC resulted in a majority of point mutations/small deletions similar to WT with the control B-DNA YAC, the Z-DNA YAC in the *Δrad1(XPF)* strain differed from the WT, resulting in an increase in small deletions within the *URA3* gene, rather than in increase in DSBs and arm loss as seen in the WT strain (Fig. 2.13c). It is clear that in these NER- and MMR-deficient strains [(*Δrad10(ERCC1)*, *Δrad1(XPF), Δmsh2,* and *Δmsh3*], frank DSBs and arm loss is markedly decreased in both control and Z-DNA-forming YACs compared to the WT, suggesting that these repair proteins are involved in DSB formation in yeast, while other proteins in similar pathways are not. Comprehensive results of mutation spectra for all strains tested are listed in Table 2.4.

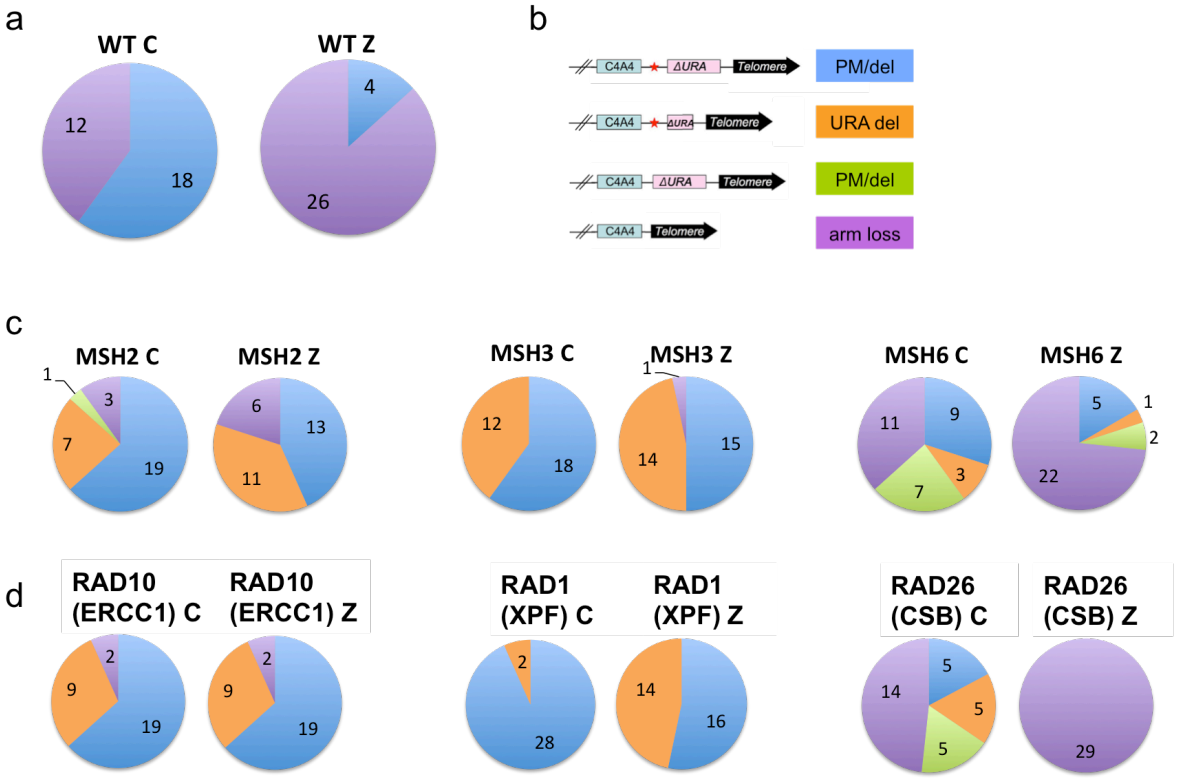


Figure 2.13. Z-DNA-induced mutation spectra in wild-type and repair-deficient yeast. FOA^R mutants were analyzed by PCR and gel electrophoresis as shown in Fig. 2.12. **(a)** In wild-type cells, the $(CG)_{14}$ insert (WT Z) resulted in greater arm loss events caused by a frank DSB (purple), while the control insert (WT C) resulted in a greater number of point mutations and small deletions within the *URA3* gene (blue). **(b)** Possible mutagenic events detected by this assay corresponding to the graphs in this figure. Mutation spectra of representative MMR-deficient **(c)** and NER-deficient **(d)** strains that varied or were similar to wild-type spectra. These experiments were performed by Jennifer McKinney.

Table 2.4. Mutation spectra of wild-type and repair-deficient mutant yeast.

Strain	YAC	% PM/del (three products present)	% URA deletion (M13 product present)	% PM/del (URA3 product present)	% arm loss (LEU2 product present)
WT (BY4742)	**C**	**60**	**0**	**0**	**40**
	Z	**13**	**0**	**0**	**87**
rad4Δ (XPC)	C	48	21	7	24
	Z	0	100	0	0
rad26Δ (CSB)	C	17	17	17	47
	Z	0	0	0	100
rad14Δ (XPA)	C	13	87	0	0
	Z	0	90	0	10
rad1Δ (XPF)	C	93	7	0	0
	Z	53	47	0	0
rad2Δ (XPG)	C	69	24	0	7
	Z	20	67	3	10
rad10Δ (ERCC1)	C	63	30	0	7
	Z	63	30	0	7
msh2Δ	C	63	23	3	10
	Z	43	37	0	20
msh3Δ	C	60	40	0	0
	Z	50	47	0	3
msh6Δ	C	30	10	23	37
	Z	17	3	7	73
mlh1Δ	C	73	0	7	20
	Z	47	3	7	40
exo1Δ	C	58	10	6	27
	Z	0	0	0	100
rad27Δ (FEN1)	C	0	17	0	83
	Z	0	0	0	100
mre11Δ	C	47	13	10	30
	Z	0	0	17	83
sae2Δ	C	43	0	10	47
	Z	0	0	7	100
rad50Δ	C	90	0	7	3
	Z	13	0	30	57

*Mutation spectra of WT in bold; mutants with spectra notable different than wild-type are denoted in red. n≥30.
[McKinney, *et al.* (submitted *Molecular Cell* 2016)]

2.4.4. Correlation of mutation frequency and spectra data as determined by mutation index (MI).

Although several mutant strains tested displayed an effect in both mutation frequency and spectra assays that differed from the wild-type, suggesting a role in Z-DNA-induced mutagenesis in yeast, a few of the strains demonstrated a more prominent effect in one of the assays and not the other. To determine which gene products we tested had the most meaningful effect on Z-DNA-induced mutagenesis, and thus the most-likely component(s) involved in the mutagenic mechanism, we needed to consider data collected from both assays in a way that did not emphasize the importance of one assay over the other.

As elaborated in Materials and Methods (Section 2.3.6.), we combined the data from mutation frequencies and the data from mutation spectra to compute MI, a single mutation index whose magnitude is directly proportional to the mutagenic potential of a Z-DNA-forming sequence in a given yeast strain. Thus, a drop in MI may be used as a proxy for determining whether a repair protein was relevant or not in eliciting Z-DNA-dependent mutations. Importantly, ranking based on decreasing MI values provides a snapshot of the relevance each repair protein played in Z-DNA-induced mutagenesis, combining data from mutation frequency and spectra assays (Fig. 2.14).

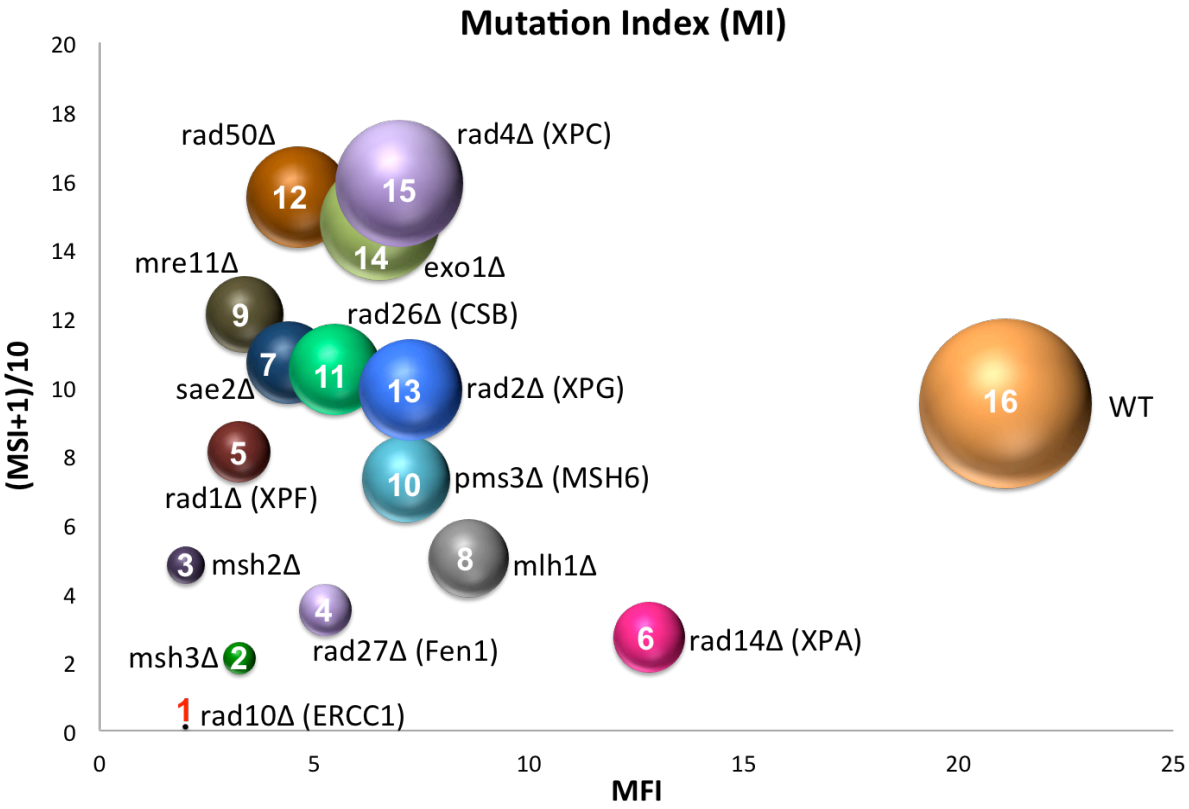


Figure 2.14. Compilation of mutation frequency and spectra data as mutation index (MI). Bubble plot of MSI versus MFI, in which the area of each circle was computed from the value of MI = MFI * (MSI + 1) / 10, as the radius. Ranking is listed as the number on each bubble with the lowest ranking representing that of the mutant strain with the greatest divergence from wild-type (i.e., *rad10Δ(ERCC1)*, ranked 1 versus wild-type, ranked 16). These data prepared by Jennifer McKinney with assistance from Dr. Albino Bacolla. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

As expected, the MI was highest for the wild-type strain. Interestingly, the individual components of both the Rad1-Rad10 *(XPF-ERCC1)* and Msh2-Msh3 complexes all rank within the top 5, while several of the other components of the NER and MMR pathways rank much higher, closer to WT. This suggests that the Rad1-Rad10 (XPF-ERCC1) and Msh2-Msh3 complexes play a role in Z-DNA-induced genetic instability outside of the traditional roles the complexes provide in the respective DNA repair pathways, and possibly in conjunction. Thus, we chose Rad1-Rad10 (XPF-ERCC1) and Msh2-Msh3 to continue our studies and further determine if there is a dependent relationship between the two complexes and/or nuclease activity that is required for Z-DNA-induced mutagenesis.

2.4.5. Rad1(XPF) and Msh3 are associated with Z-DNA-forming sequences in yeast.

To determine if the Rad10-Rad1(ERCC1-XPF) and Msh2-Msh3 repair complexes were enriched at the Z-DNA-forming sequence, and further confirm our hypothesis of these proteins in Z-DNA-induced mutagenesis, we performed ChIP assays in wild-type yeast. Using antibodies against XPF (Rad1) and Msh3, our results demonstrated a significant enrichment (3.0- and 3.3-fold, respectively) at the Z-DNA-forming insert compared to the control B-DNA-forming insert (Fig 2.15). Due to lack of availability and quality of antibodies to yeast-specific repair proteins, we were unable to test the full spectrum of repair proteins in yeast.

Furthermore, experimental difficulties including growth rate of repair-deficient strains proved difficult to perform ChIP assays on these strains. However, we were able to achieve testing of a more comprehensive panel of NER and MMR repair proteins during follow-up experiments in human wild-type and repair-deficient cells (see section 2.4.8.).

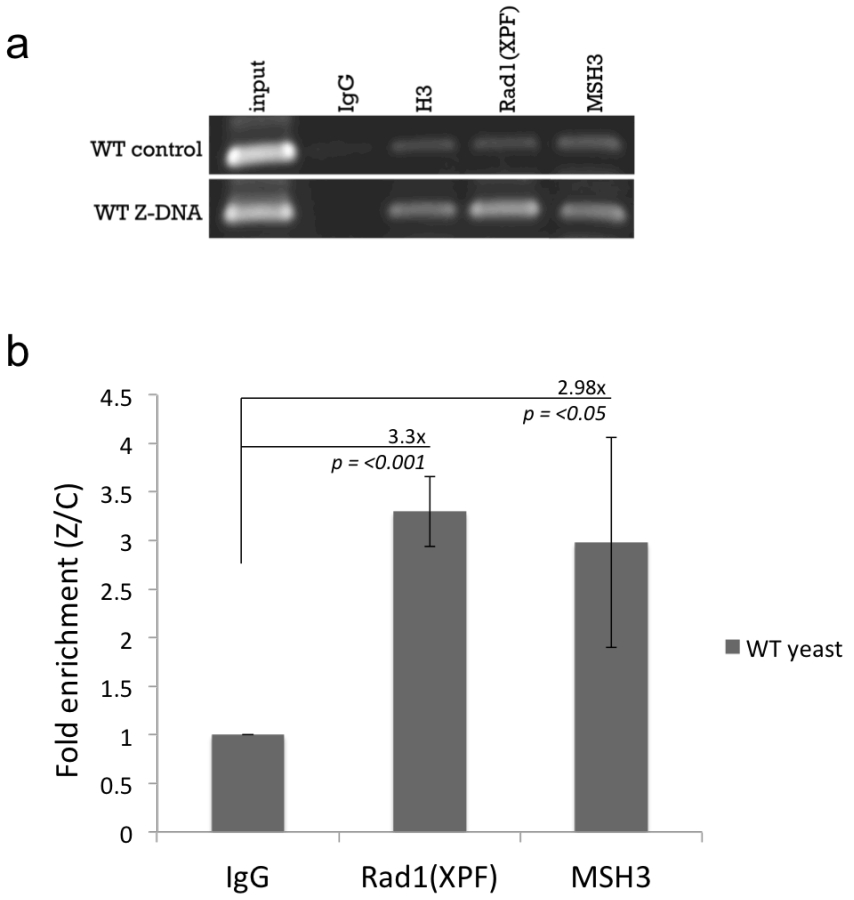


Figure 2.15. NER and MMR repair proteins are enriched at Z-DNA-forming regions on YACs in yeast. ChIP assays performed on YACs containing control B-DNA-forming (C) or Z-DNA-forming (Z) sequences in yeast wild-type BY4742 strain. **(a)** Antibodies to Rad1(XPF), Msh3, H3 (positive control) and IgG (antibody specificity control) were used for pull-down. Resulting purified chromatin and 2% input were used for PCR template and analyzed on a 1% agarose gel stained with ethidium bromide. Primer sets included M13-20 and M13Rev (Table 2.1). **(b)** Quantification of the gel revealed a 3.3-fold increase in Rad1(XPF) and ~3-fold increase in Msh3 enrichment at the Z-DNA-forming sequence when compared to the control B-DNA-forming sequence (*P* values calculated by Students *t*-test). Rad1(XPF) and Msh3 values were normalized to corresponding input, then IgG values. Experiments were performed in triplicate. These experiments were performed by Jennifer McKinney.

2.4.6. Nucleosome positioning is effected by Z-DNA in yeast.

The formation of Z-DNA has been shown to effect nucleosome positioning (59,274,275). Due to the distorted and left-handed nature of the structure, DNA that is in the Z conformation cannot wrap around histone proteins to form nucleosomes. By using a yeast-specific histone H3 antibody and performing a ChIP assay, we demonstrated that the formation of Z-DNA, which is not dependent on transcription in our system, excludes histones, thereby effecting nucleosome positioning. We saw a 13-fold decrease in H3 enrichment at the Z-DNA site when compared to the control sequence (Fig. 2.16). Our methods and results are in agreement with those published previously (59,274,275), thus we conclude that these results indirectly verify that our Z-DNA-forming sequence is forming structure on the YAC in yeast cells in our system.

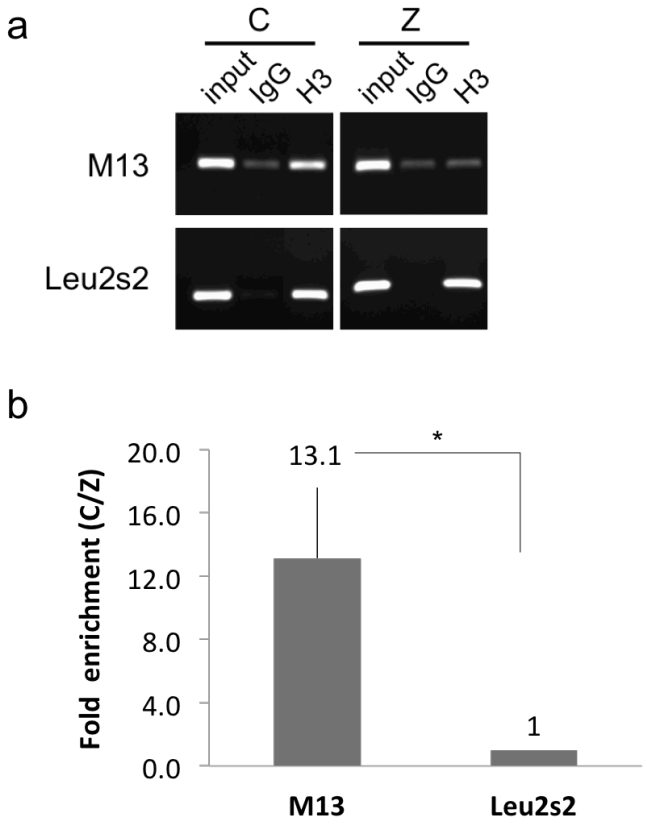


Figure 2.16. Z-DNA effects nucleosome positioning in yeast. ChIP assay performed on YAC containing control B-DNA-forming (C) or Z-DNA-forming (Z) sequences in yeast wild-type BY4742 strain. **(a)** Antibodies to H3 as well as IgG (negative control) were used for pull-down. Resulting purified chromatin and 2% input were used for PCR and analyzed on a 1% agarose gel stained with ethidium bromide. Primer sets included M13-20 and M13Rev (proximal), as well as Leu2s2For and Leu2s2Rev (distal) (Table 2.1). **(b)** Quantification of the gel revealed a 13-fold decrease in H3 enrichment at the Z-DNA-forming sequence when compared to the control B-DNA-forming sequence (*P* value <0.05 (*). H3 values with the M13 primers were normalized to H3 values with the Leu2s2 primers. Experiments were performed in triplicate. These experiments were performed by Jennifer McKinney.

2.4.7. Z-DNA-induced genetic instability in wild-type and repair-deficient human cells.

To further confirm our results obtained in yeast and to determine if the ERCC1-XPF and MSH2-MSH3 repair complexes were required for Z-DNA-induced mutagenesis in human cells, we performed mutagenesis assays via blue/white screening using control B-DNA-containing and Z-DNA-containing mutation-reporter plasmids as previously described (Fig. 2.17a) (127,262). Isogenic human XPF-proficient (GM08437B-XPF) (wild-type) or XPF-deficient (GM08437B-pLPC) cells (Fig. 2.17b, top panel) were transfected with mutation-reporter plasmids containing either a control B-DNA-forming or Z-DNA-forming sequence (Fig. 2.17a). Using blue/white screening mutagenesis assays, we found that the Z-DNA-forming plasmid resulted in a significant 6.7-fold increase in mutations in wild-type human cells when compared to the control plasmid (Fig. 2.18, purple bars, *P* value $1x10^{-5}$). In contrast, we found that in the XPF-deficient human cells, the Z-DNA-forming plasmid resulted in only a 1.1-fold increase when compared to the control (Fig. 2.18, green bars, *P* value 0.11), suggesting that the ERCC1-XPF repair complex is required for the significant increase in Z-DNA-induced mutagenesis seen in wild-type cells. These results are in accordance with our yeast data.

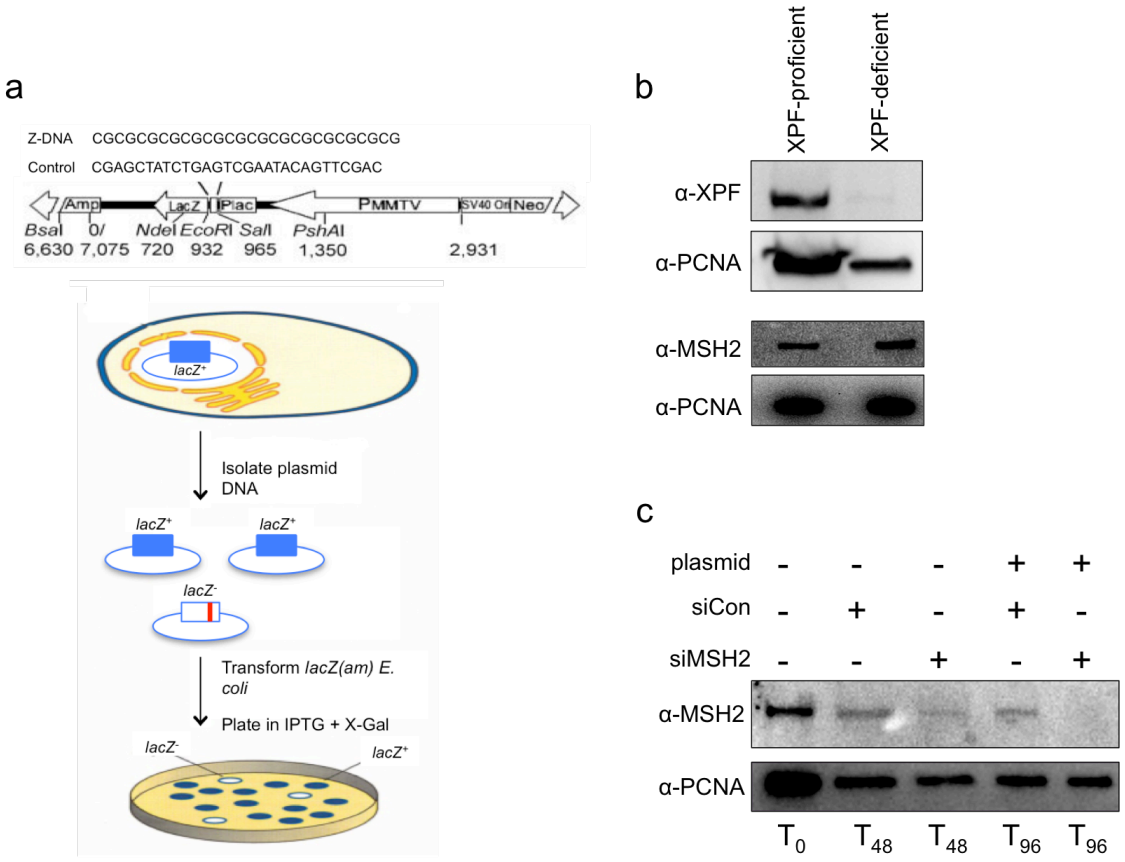


Figure 2.17. Cell lines used for mutagenesis assays in human cells. Blue/white screening using mutation-reporter plasmids was performed to determine levels of Z-DNA-induced mutagenesis in human cells. **(a)** Schematic of mutation-reporter plasmid (variation of pUCNIM, 7075 bp) containing control B-DNA-forming or Z-DNA-forming sequence [$(CG)_{14}$] inserted between the *LacZ* gene and promoter [top panel, adapted from (127)]. Reporter plasmids (*LacZ*$^+$) were transfected into human cells, extracted and purified after 48 hours, transformed into bacterial cells, and plated onto media containing X-Gal to screen for mutants (white colonies, *LacZ*$^-$) as a frequency compared to wild-type (blue colonies, *LacZ*$^+$) [bottom panel, adapted from (276)]. **(b)** Human XPF-proficient and -deficient cell lines confirmed by Western blotting using antibodies against XPF (top panel) and MSH2 (bottom panel). PCNA was used as a loading control. **(c)** MSH2 depletion by siRNA knockdown in human XPF-proficient cells confirmed by Western blotting using an antibody against MSH2, and PCNA as a loading control. Plasmid only and a non-targeting siRNA (siCon) were used as negative controls. Various time points were tested, indicating hours following first transfection. These experiments were performed by Jennifer McKinney. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

Similarly, we wanted to determine if the MSH2-MSH3 complex was required for Z-DNA-induced mutagenesis. Using targeted siRNA knockdown, we depleted MSH2 from the XPF-proficient cells (wild-type, and thus are also MSH2-proficient, Fig. 2.17b, bottom panel) resulting in 80-95% knockdown (Fig. 2.17c). Following treatment with a non-targeting siRNA, the Z-DNA-forming plasmid stimulated an ~5-fold increase in mutations compared to the control B-DNA-forming plasmid (Fig. 2.18, blue bars, *P* value 0.002), similar to the induction seen in wild-type cells. Depletion of MSH2 by siRNA knockdown in the wild-type cells resulted in a decrease of Z-DNA-induced mutagenesis to 1.1-fold over control (Fig. 2.18, red bars, *P* value 0.78), suggesting that MSH2 was required for the Z-DNA-induced mutagenesis seen in wild-type cells. Direct DNA sequencing was performed on mutant plasmids from human wild-type, XPF-deficient, and MSH2-depleted cell lines to determine mutation spectra; however no difference was seen in the types of mutations in either plasmid across cell lines. The majority of mutations in all cases were large deletions, with no obvious bias in location relative to the control B-DNA-forming or Z-DNA-forming inserts. This may be explained by the tendency of this particular pUCNIM plasmid to detect large-scale deletions. Nonetheless, while the types of mutation did not differ significantly, the frequency of these mutations was significantly higher with the Z-DNA-forming plasmid when compared to control.

The results obtained in human cells were consistent with our yeast data suggesting that repair complexes from both NER and MMR pathways are

involved in Z-DNA-induced mutations, possibly acting in a novel pathway outside of their canonical roles within the respective DNA repair pathways. Previous results from our laboratory have shown that there is no significant effect on Z-DNA-induced mutagenesis in human XPA-deficient cells when compared to wild-type cells (unpublished data), suggesting that functional NER was not required for the Z-DNA-induced mutagenesis. Thus, we continued our focus on the ERCC1-XPF and MSH2-MSH3 repair complexes to further characterize the mechanism responsible for Z-DNA-induced genomic instability.

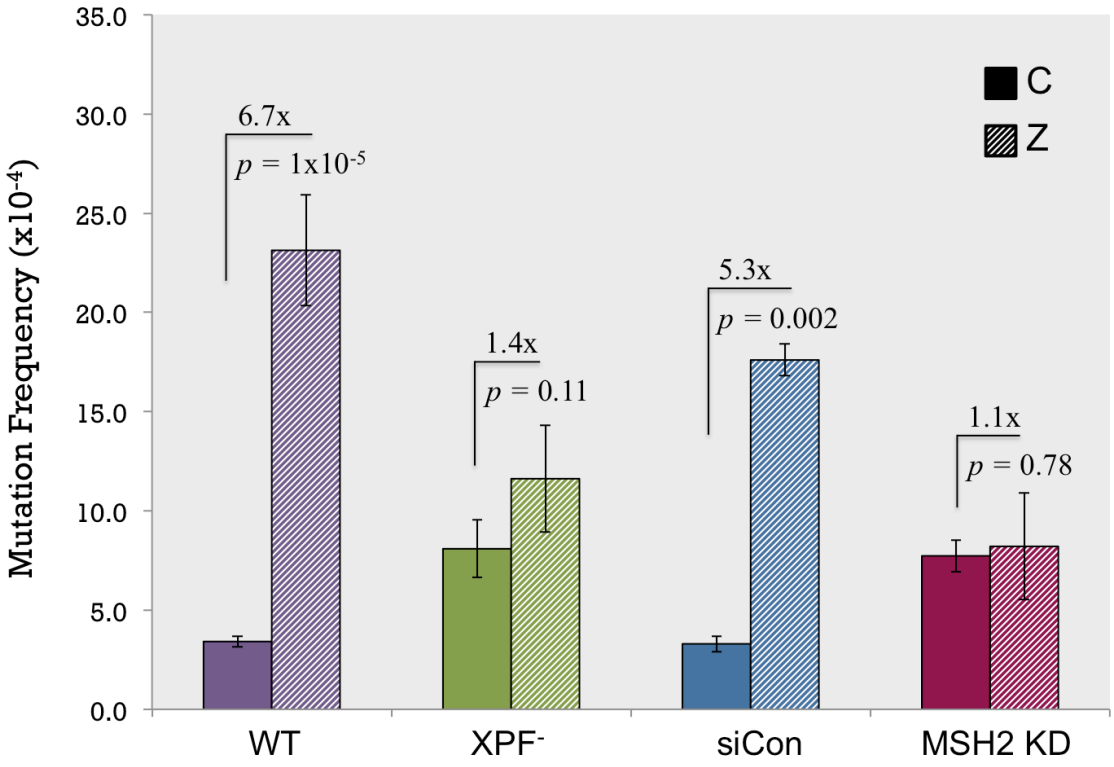


Figure 2.18. Z-DNA-induced mutagenesis is decreased in human XPF-deficient and MSH2-deficient cells. Plasmids containing control B-DNA-forming or Z-DNA-forming sequences were transfected into human wild-type and repair-deficient human cell lines. Plasmids were isolated 48 hours following transfection and subjected to bacterial blue/white screening to determine mutation frequencies (Fig. 2.16a). Mutation frequencies were calculated as an average from three separate experiments in wild-type (purple bars), XPF-deficient (green bars), wild-type treated with sicontrol (blue bars), and siRNA-depleted MSH2 (red bars) cell lines. Statistical differences between Z-DNA and control B-DNA plasmids within cell lines (diagonal lines *vs.* solid bars, respectively) were calculated using Student's *t*-test. The Z-DNA-forming plasmid resulted in 6.7-fold increase in mutations, which was significantly decreased following depletion of either XPF or MSH2. These experiments were performed by Jennifer McKinney. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

2.4.8. Association of NER and MMR proteins with Z-DNA-forming sequences in human cells.

To further investigate the involvement of ERCC1-XPF and MSH2-MSH3 repair complexes in Z-DNA-induced mutagenesis, we tested whether these complexes were enriched at a Z-DNA-forming sequence in human cells. We performed ChIP assays in the wild-type and repair-deficient human cell lines that were used for the mutagenesis studies. Using antibodies against NER (XPA and XPF) and MMR proteins (MSH2, MSH3, MSH6), our results in wild-type and XPF-deficient cells demonstrated an enrichment at the Z-DNA-forming insert compared to the control B-DNA-forming insert of XPF (4-fold in wild-type cells only, as XPF-deficient cells cannot be considered, because, in theory, there is no functional XPF in these cell lines) (Fig. 2.19a and d), MSH2 and MSH3 (5-6-fold, and 4-9-fold, respectively) (Fig. 2.19b and d). This result suggests that these proteins have a higher affinity for Z-DNA than B-DNA, perhaps playing roles to recognize and process the structure, and that XPF is not required for recruitment of MSH2-MSH3 to the Z-DNA-forming insert, as it is enriched at Z-DNA in both XPF-proficient and XPF-deficient cell lines. Moreover, in both cell lines, XPA demonstrated only a slight enrichment (1.8-2-fold) at the Z-DNA-forming insert over control, and MSH6 exhibited no fold change. When comparing enrichment between cell lines, the values did not change between XPF-proficient and XPF-deficient cell lines, other than the expected enrichment of XPF, suggesting that

the lack of XPF does not influence the enrichment of MSH2-MSH3 (Fig. 2.19a and b, values represented below each gel).

To determine if ERCC1-XPF recruitment to the Z-DNA-forming insert was dependent on MSH2-MSH3, we performed ChIP assays on MSH2 siRNA knockdown human cell lines. Interestingly, the knockdown of MSH2 resulted in the loss of the enrichment of XPF at the site of the Z-DNA-forming insert that was present in wild-type (XPF-proficient) cells (4-fold enrichment of Z/C in wild-type decreased to 1.1-fold in MSH2 KD cells) (Fig. 2.19c and d). There was no effect on XPA or MSH6 recruitment by MSH2 depletion. Taken together, these results further implicate a role for ERCC1-XPF and MSH2-MSH3 complexes in Z-DNA-induced mutagenesis as a separate component from NER and MMR pathways, rather than one or both pathways acting as a whole, and that ERCC1-XPF recruitment to a Z-DNA-forming site was dependent on MSH2-MSH3 via a novel relationship between the two repair complexes outside respective canonical pathways.

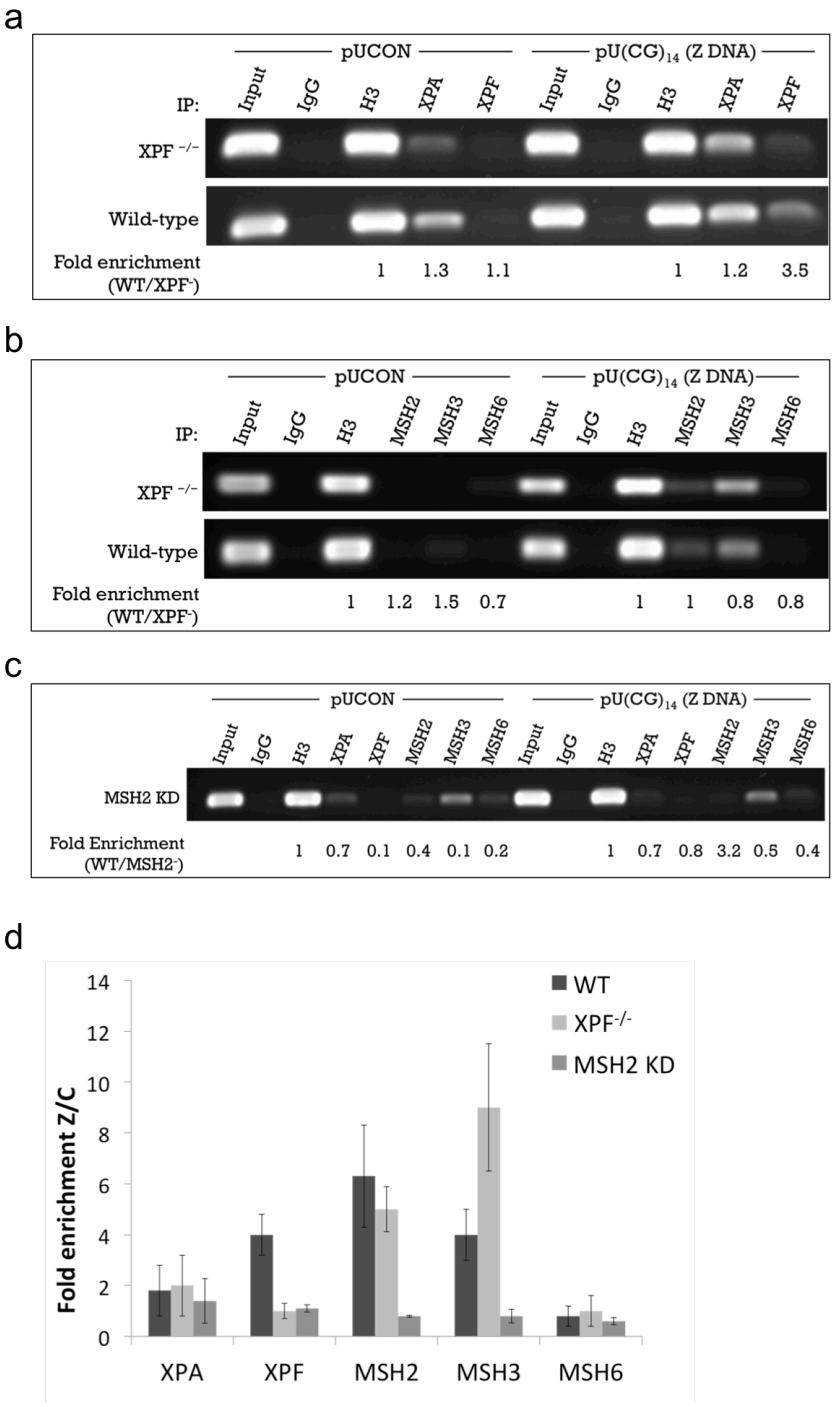


Figure 2.19. MSH2-MSH3 is required for ERCC1-XPF enrichment at Z-DNA-forming regions in human cells.

Figure 2.19. MSH2-MSH3 is required for ERCC1-XPF enrichment at Z-DNA-forming regions in human cells. ChIP analysis performed on plasmids in human cells verified enrichment of ERCC1-XPF **(a)** MSH2/MSH3 **(b)** repair complexes at a Z-DNA-forming sequence ($pU(CG)_{14}$) when compared to control (pUCON) in wild-type (WT) and XPF-deficient human cell lines ($XPF^{-/-}$). No significant enrichment of XPA or MSH6 was detected. **(c)** In MSH2 depleted cells by siRNA-knockdown (MSH2 KD), the enrichment of XPF at Z-DNA was not detected. Antibodies to IgG and H3 were used as negative and positive controls, respectively. **(d)** Quantification of enrichment comparing the Z-DNA-forming region to the control B-DNA-forming region (Z/C). All values were an average of three repeats, normalized to input. Fractions of purified ChIP and input DNA were used for PCR analysis. Primers used for PCR amplification included pUinsFor1 and pUinsRev1 (Table 2.1). Amplified products were analyzed on 1% agarose gels stained with ethidium bromide. These experiments were performed by Jennifer McKinney and Dr. Anirban Mukherjee. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

2.4.9. In vitro cleavage of Z-DNA by the NER nuclease complex, ERCC1-XPF.

To determine if ERCC1-XPF can process Z-DNA, we first confirmed Z-DNA structure formation on plasmids *in vitro*. When Z-DNA formation occurs, single-stranded regions are formed at the B-Z junctions (18,22), and S1 nuclease assays have been widely used to assess the presence of Z-DNA and other non-B DNA structures (125,127,134,277). We treated the control B-DNA (pUCon) and Z-DNA-forming plasmid [pU(CG_{14})] with S1 nuclease, followed by restriction digestion to form fragments of known lengths and to detect fragments caused by S1 nuclease to confirm Z-DNA formation (Fig. 2.20a). Following S1 treatment, the plasmid with the Z-DNA-forming sequence clearly resulted in a cleaved fragment mapping to the Z-DNA-forming insert that was not present with the control plasmid (Fig. 2.20b). These results confirm the presence of Z-DNA on the pU$(CG)_{14}$ plasmid, providing a viable substrate to determine if ERCC1-XPF could process Z-DNA.

Previously, we have shown that Z-DNA-forming sequences can stimulate the formation of DSBs, resulting in large-scale deletions (127) and in this study we have shown that ERCC1-XPF is likely to be involved in this mechanism. Initially, we performed linker-mediated PCR (LM-PCR) using control B-DNA and Z-DNA-forming plasmids in isogenic human XPF-proficient and XPF-deficient cells.

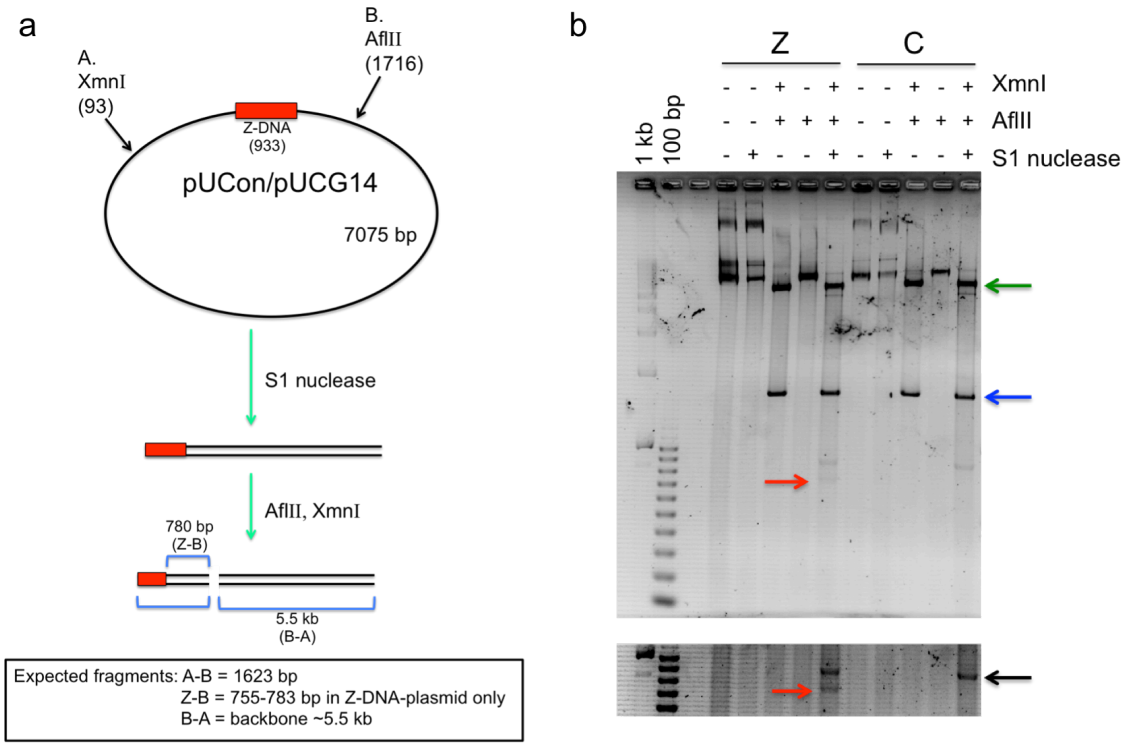


Figure 2.20. Formation of Z-DNA on plasmid substrates as detected by S1 nuclease sensitivity. (a) S1 nuclease assay schematic. Plasmid DNA containing control B-DNA-forming sequence (pUCon) or Z-DNA-forming sequence [pU(CG)$_{14}$] (red box) was treated with S1 nuclease, then digested with restriction endonucleases XmnI (A) and AflII (B) to release the fragment containing the insert (red box). Fragments were separated on 1.4% agarose gels stained with SYBR® Gold and visualized on a ChemiDoc Imaging System. S1 nuclease cleaves at single-stranded regions of DNA. If Z-DNA formation occurs, then several fragments are expected as shown above. The control plasmid should yield fewer fragments, as it does not form Z-DNA. **(b)** Top panel: SYBR Gold-stained agarose gel demonstrating Z-DNA structure formation as determined by an extra fragment at ~700 bp (red arrows) in the pU(CG)$_{14}$ plasmid that is absent in the pUCon plasmid. Bottom panel: enlarged and higher exposure of 700-bp fragment. Blue arrow represents A-B fragment; green arrow: represents plasmid backbone. Black arrow represents a fragment resulting from a secondary structure that forms a substrate for S1 nuclease between the XmnI restriction site and the insert on both Z-DNA and control plasmids. Experiments were performed in triplicate to confirm results. These experiments were performed by Jennifer McKinney.

Our results clearly demonstrate that DSBs are indeed formed at the Z-DNA-forming region as indicated by the presence of a major band at 210 bp, revealing a hotspot for breakage in the $(CG)_{14}$ sequence, which is absent at this location in the control B-DNA plasmid (Fig. 2.21). However, similar to our mutation spectra results, DSBs were detected both in the presence and absence of XPF and we were unable to quantify a difference between the proficient and deficient cell lines. These studies are on-going to optimize conditions to determine if a quantifiable difference exits between cell lines.

ERCC1-XPF nicks DNA on the 5' side of a lesion; however, it is unknown whether ERCC1-XPF alone can lead to DSBs via multiple nicks, or if ERCC1-XPF nicks once and another component is responsible for subsequent processing leading to DSBs. Thus, we developed an assay to detect single-stranded nicks on either side of the Z-DNA-forming sequence. We performed *in vitro* cleavage assays using whole cell extract (WCE) from the human XPF-proficient (WT) and XPF-deficient cell lines used in the mutagenesis and ChIP assays with the control B-DNA-forming or Z-DNA-forming plasmid. The addition of aphidicolin was used to block the repair synthesis of nicks or breaks that occurred, allowing easier detection of the nicks (Fig. 2.22a). By using PCR to extend a primer up or downstream of the insert, we detected shorter products in the samples containing the Z-DNA-forming sequence that were incubated with XPF-proficient WCE that were not present in either the control samples or the XPF-deficient WCE with the Z-DNA-forming sequence (Fig. 2.22b, red arrows).

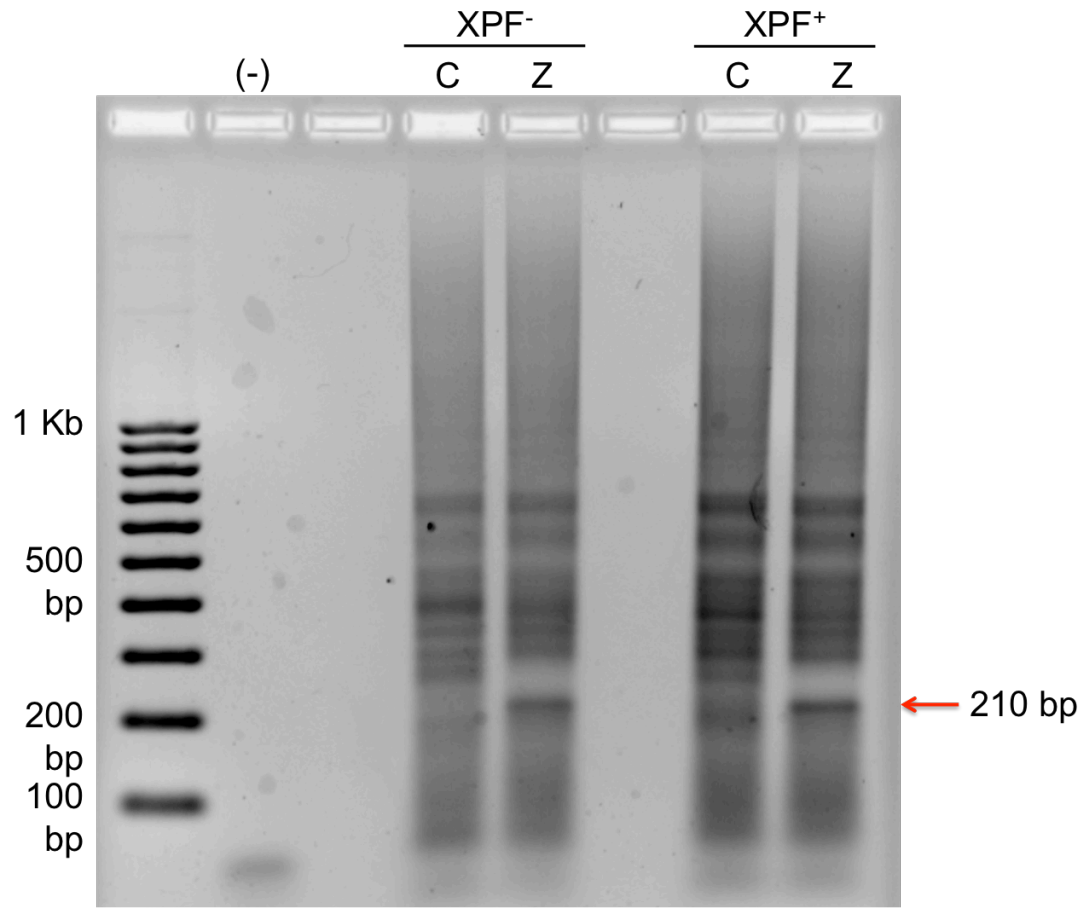


Figure 2.21. Analysis of Z-DNA-induced DSBs by LM-PCR. Control B-DNA and pU(CG)$_{14}$ Z-DNA plasmids were transfected into isogenic human XPF-proficient (XPF^{+}) and XPF-deficient (XPF^{-}) cells. Plasmids were purified and used as substrates to perform LM-PCR. Final LM-PCR products were separated on a 1.2% agarose gel, stained by ethidium bromide and visualized. Bands indicate breakpoints, and a major breakpoint at the location of the Z-DNA-forming region on the pU(CG)$_{14}$ Z-DNA plasmid (Z) resulted in a prominent band at 210 bp (red arrow) that is not detected at the same location in the control B-DNA control (C) plasmid. A negative control for the PCR reaction was used in which no template was added (-). This experiment was performed by Jennifer McKinney and Dr. Guliang Wang.

Interestingly, shorter products were detected with both the left and right primers suggesting that ERCC1-XPF may cleave both upstream and downstream on the 3' side of the Z-DNA-forming sequence.

A frank DSB at the site of insert would result in a product ~160-180 bp in length, similar to the EcoRI positive control used in this assay (Fig. 2.22b, black arrow). Although the extension of the primers terminated ~500-600 bp away from the insert, this distance may not reflect the exact site of cleavage. Aphidicolin prevents DNA synthesis by blocking DNA polymerase α (278,279), therefore it is likely that the 500-600 bp products are a result of subsequent nuclease activity, resulting in chewing back of the DNA from the ERCC1-XPF-generated nick on either side of the insert during repair of the "damage". The data from this result suggests that ERCC1-XPF can cleave at a Z-DNA-forming site further implicating this complex in Z-DNA-induced mutagenesis in eukaryotes.

The results from this study suggest a novel mechanism for Z-DNA-induced mutagenesis in eukaryotes involving components of DNA repair working outside of traditional repair pathways. We propose a model (Fig. 2.23) in which Z-DNA formation occurs during various DNA metabolic processes (i.e., transcription, replication, etc.), which is recognized by MSH2-MSH3 (and other possible factors) as "damage" and triggers a DNA damage response. MSH2-MSH3 recruits ERCC1-XPF (and other possible factors) to the site of "damage" (Z-DNA structure), which processes the DNA in an attempt to remove the "damage". This processing, if error free, can remain undetected with no adverse

outcomes. However, if error-prone, DSBs can occur, potentially resulting in large deletions or translocations at or near the site of “damage”, possibly explaining how breaks and translocations map to sites of Z-DNA-forming sequences (122,123,161,166,258).

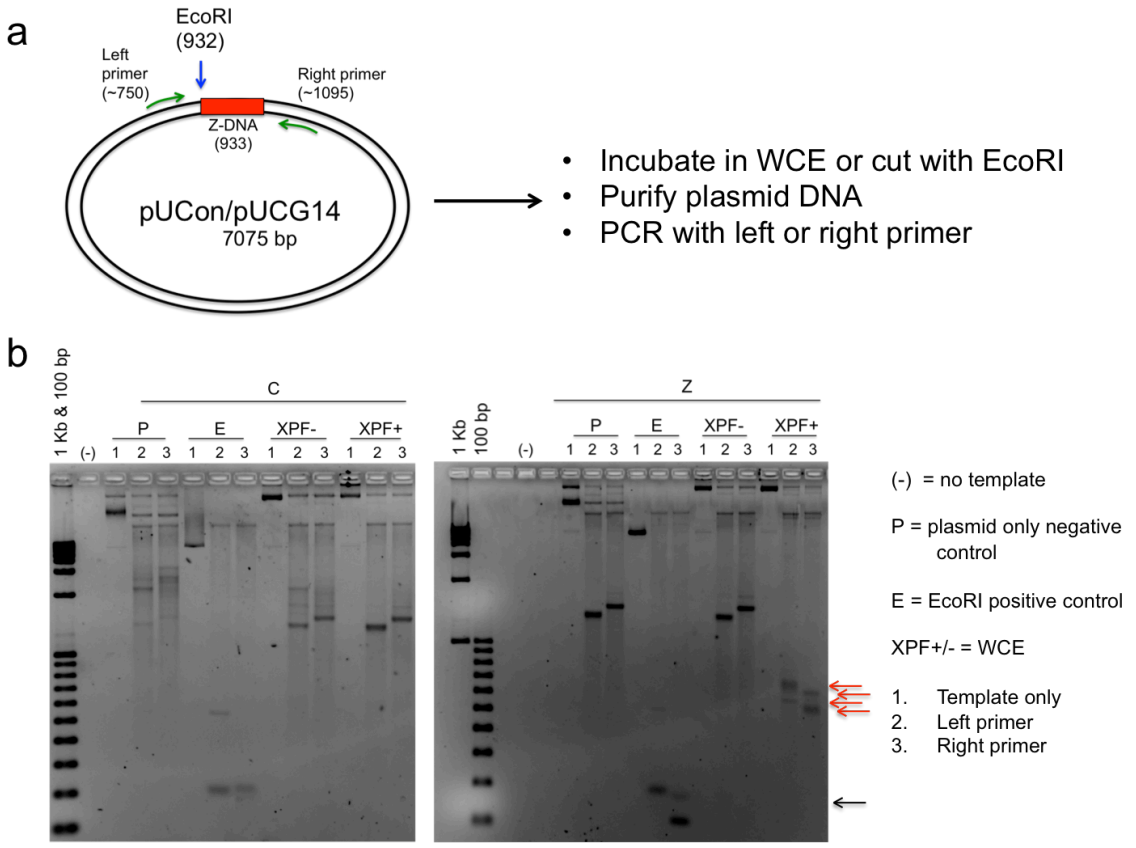


Figure 2.22. ERCC1-XPF cleaves near Z-DNA-forming sequences in plasmid DNA. (a) Schematic of assay. Plasmid DNA containing a B-DNA-forming control (C) or a Z-DNA-forming sequence (Z) was incubated in human XPF-proficient or XPF-deficient whole cell extract (WCE) containing aphidicolin to block repair of nicks, or digested with EcoRI. DNA was purified and used as a template for a PCR reaction using either the left or right primer (green arrows). PCR products were separated on 1.5% agarose gels stained with SYBR® Gold and visualized on a ChemiDoc Imaging System. **(b)** Representative gel from PCR primer extension assay demonstrates that cleavage of the area surrounding the Z-DNA-forming insert on plasmid DNA occurs in XPF-proficient WCE that does not occur in the control plasmid or in XPF-deficient WCE as seen by extra cleavage products (red arrows). Included in each experiment were a negative control of plasmid DNA only, as well as a positive control in which plasmids were restricted with EcoRI at the site of the insert to demonstrate that the assay would detect a break resulting in a ~160 or ~180 bp product with the right or left primers, respectively (black arrow). These experiments were performed by Jennifer McKinney. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

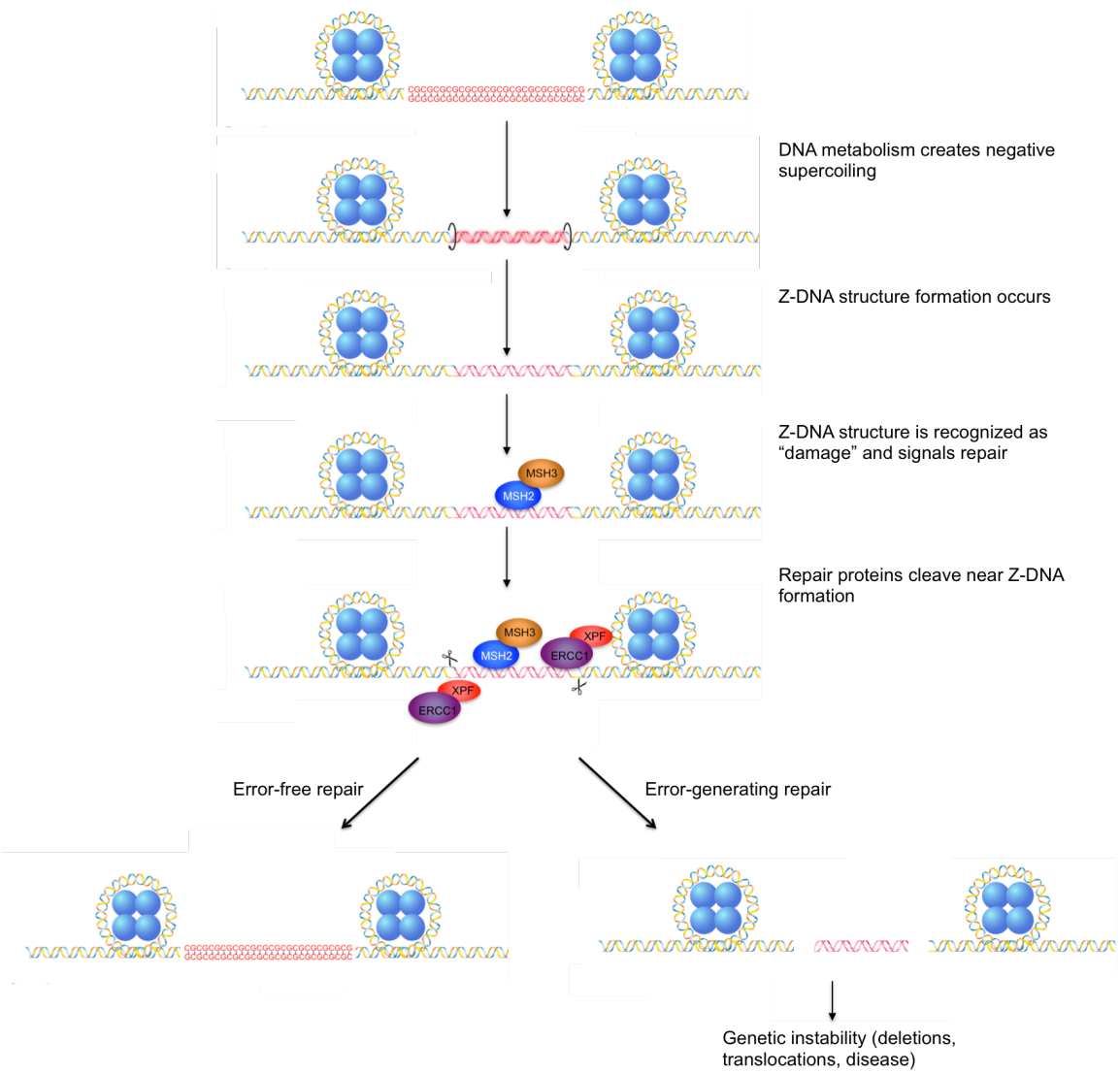


Figure 2.23. Model for Z-DNA-induced mutagenesis. During DNA metabolic processes such as replication, transcription, repair, etc., negative supercoiling is generated, stimulating Z-DNA formation. The structure of Z-DNA is recognized as "damage" by the repair complex, MSH2-MSH3 (and other possible factors), signaling repair. The ERCC1-XPF complex (and other possible factors) is recruited to the site for cleavage near the Z-DNA-forming region, resulting in DSBs in an attempt to repair the "damage". The breaks may be repaired by factors yet to be identified, resulting in error-free repair, or the breaks may be processed by an end-joining mechanism resulting in genomic instability in the form of large deletions, translocations, which may contribute to disease etiology. This model was prepared by Jennifer McKinney with the assistance of Dr. Karen Vasquez. [McKinney, *et al.* (submitted *Molecular Cell* 2016)].

2.5. Discussion

We have shown that Z-DNA-forming sequences can lead to increased genetic instability and DSBs in various model systems including bacteria, mammalian cells, and mice (60,127,128,132,171). Importantly, we, and others, have shown that sequences that can form Z-DNA have been shown to map to translocation breakpoint hotspots associated with leukemias and lymphomas (122,123,161,166,258). Here, we identified two major components involved in a mechanism responsible for Z-DNA-induced mutagenesis in eukaryotes.

To our knowledge, this is the first demonstration that Z-DNA is mutagenic in yeast. We have shown that a $(CG)_{14}$ Z-DNA-forming sequence is significantly more mutagenic in wild-type yeast strains, and leads to frank DSBs near the Z-DNA-forming sequence, when compared to a control B-DNA-forming sequence. By using a yeast artificial chromosome (YAC) containing a control B-DNA or Z-DNA-forming sequence to screen a deletion library of repair-deficient yeast strains, we identified several gene products that are likely to be involved in a mechanism of Z-DNA-induced mutagenesis. Interestingly, the most striking of these gene products (i.e., the gene deficiency resulting in the greatest loss of effect seen in wild-type caused by Z-DNA) included the NER and MMR repair complexes Rad10-Rad1(ERCC1-XPF) and Msh2-Msh3, respectively. In yeast cells deficient in Msh2, Msh3, Rad10(ERCC1), or Rad1(XPF), the frequencies and types of mutations seen in the Z-DNA-containing YAC were significantly

reduced to background levels when compared to wild-type cells. Thus, in the absence of these repair complexes, a Z-DNA-forming sequence was no longer mutagenic, suggesting that the Msh2-Msh3 and Rad10-Rad1 complexes play an important role in Z-DNA-induced genetic instability in yeast.

We confirmed our results from yeast in wild-type and repair-deficient human cells using a plasmid-based mutagenesis assay (127,262). A plasmid containing a Z-DNA-forming sequence resulted in significantly higher mutations in wild-type human cells when compared to a control plasmid that did not contain Z-DNA-forming sequences. However, deficiency in XPF or MSH2 resulted in a decrease of mutations to near background levels when compared to control, wild-type cells, similar to our results in yeast. Furthermore, we demonstrated that both ERCC1-XPF and MSH2-MSH3 complexes, but not other repair proteins in the NER or MMR mechanisms (i.e., XPA and MSH6, respectively), were enriched at a Z-DNA-forming sequence, and that ERCC1-XPF enrichment to the Z-DNA-forming region was dependent on MSH2-MSH3, suggesting a relationship between the two complexes outside of canonical NER and MMR, respectively. We speculate that MSH2-MSH3 may recognize and bind to Z-DNA-forming sequences and recruit ERCC1-XPF for processing of the “damage” (Fig. 2.22).

Although we were able to demonstrate the formation of DSBs at the Z-DNA-forming sequence on plasmids in human cells by LM-PCR, we were unable to detect a quantifiable difference between WT and XPF-deficient cells lines. The LM-PCR results support the mutation spectra data we obtained in human cells

using the plasmid-based mutagenesis assay in which we saw mostly large deletions in WT and XPF-deficient cell lines; however, the LM-PCR assay may not have been sensitive enough to detect differences between cell lines. By using an *in vitro* assay with whole cell extract from wild-type and XPF-deficient human cells we found that wild-type WCE resulted in cleavage products near the Z-DNA-forming site that were not present in the non-Z-DNA-forming site or in XPF-deficient WCE, suggesting that ERCC1-XPF can cleave on either side of the Z-DNA-forming sequence, identifying a new substrate for this structure-specific nuclease.

Collectively, the data obtained from our yeast screen, as well as from our studies in human cells, suggest that other NER and MMR proteins are not likely to be involved in Z-DNA-induced mutagenesis in yeast, implicating a *novel* function for Msh2-Msh3 and ERCC1-XPF(Rad10-Rad1) complexes acting outside of the canonical DNA repair pathways which these protein complexes are known to participate. Msh2-Msh3 and Rad10-Rad1 have been shown to cooperate in yeast for functions other than MMR and NER pathways, including processing intermediates leading to efficient recombination in targeted gene replacement (TGR) and single-strand annealing (SSA) pathways, suggesting the role of Msh2-Msh3 is to either recruit or aid in the cleavage activity of Rad10-Rad1 (280-291). Msh2-Msh3 and Rad10-Rad1 are also known to cooperate in the resolution of mismatches that can form between heteroalleles during meiotic recombination (292-295). However, neither Msh6, nor other MMR factors, were

found to be required for SSA or loop repair in meiotic recombination (282,283,287,291,293).

In mammalian cells, ERCC-XPF and MSH2-MSH3 have also been shown to interact outside of the NER and MMR pathways. For example, ERCC1-XPF and MSH2-MSH3 have been shown to cooperate in the mechanism responsible for cellular resistance to chemotherapeutic interstrand crosslinking (ICL) agents such as *cis*-diamminedichloroplatinum(II) (CDDP) by facilitating ICL repair (296). Additionally, MSH2 and ERCC1-XPF (in addition to other repair factors) have been shown to interact in the repair of site-directed psoralen-induced ICLs (297,298). A physical interaction between MSH2 and Rad1(XPF) and Rad10(ERCC1) has been shown in yeast by two-hybrid analysis (291) as well as in mammalian cells (296). Previously, our lab has shown that MMR components MSH2-MSH3 and MLH1 are required for the repair of triplex-forming oligonucleotide (TFO)-directed psoralen ICLs (252,253), and that MMR and NER proteins can interact in recognizing TFO-directed ICLs (130), further implicating roles for such proteins outside of the canonical MMR and NER pathways during DNA damage recognition and processing.

We have identified MSH2-MSH3 and ERCC1-XPF complexes to be involved in the mechanism responsible for Z-DNA-induced mutagenesis in eukaryotes. It is likely that other components are involved in this mechanism and subsequent studies are warranted to identify candidates, which will provide insight into the contribution of Z-DNA-induced genomic instability to the

development of human disease. Understanding the mechanisms involved in DNA structure-induced mutagenesis in disease etiology will allow for the advancement of therapeutics to treat and/or prevent diseases of genetic instability, such as cancer.

Chapter 3: The Role of DNA Repair Proteins in H-DNA-induced Mutagenesis in Eukaryotes[2]

[2] Sections of this chapter are included in the following publications:

Junhua Zhao, Guliang Wang, Imee M. del Mundo, **Jennifer A. McKinney**, Albino Bacolla, Stephen B. Boulware, Haihua Zhang, Catherine H. Freudenreich, and Karen M. Vasquez. Distinct mechanisms of nuclease-directed DNA structure-induced genetic instability, *Nature Communications* (submitted) 2016. For this manuscript, I was involved in executing experiments, and analysis of data.

3.1. Abstract

Repetitive DNA sequences capable of adopting alternative structures (i.e. H-DNA) co-localize with chromosomal breakage/mutation hotspots and induce genetic instability. Previously, we found that an H-DNA-forming sequence from the *c-MYC* promoter that maps to a translocation hotspot in cancer, stimulates genetic instability in mammalian cells and mice, implicating H-DNA-forming sequences in disease etiology. However, the mechanism(s) of H-DNA-induced genetic instability is unknown. We report that H-DNA-forming sequences are enriched at translocation breakpoints in human cancer, implicating H-DNA in cancer etiology. We demonstrate *for the first time* that a 23-bp H-DNA-forming sequence is mutagenic in wild-type *Saccharomyces cerevisiae* cells when compared to a control sequence. We screened repair-deficient cells and found that H-DNA-induced mutagenesis was suppressed in yeast and human cells deficient in the nucleotide excision repair nucleases, ERCC1-XPF and XPG; and in contrast, was stimulated in cells deficient in FEN1, a replication-related endonuclease. We identified novel mechanisms with H-DNA as a substrate for cleavage by ERCC1-XPF, XPG, and FEN1 in distinct replication-independent and replication-dependent pathways of genetic instability. These results provide critical information for understanding the etiology of cancer and other DNA structure-related human diseases.

3.2. Introduction

Repetitive DNA sequences capable of adopting alternative DNA secondary structures (i.e. non-B DNA) are abundant in the human genome, and often co-localize with endogenous disease-related mutation and breakpoint hotspots (11,299,300). For example, H-DNA-forming sequences are found at translocation hotspots in the *c-MYC* gene in Burkitt's lymphoma, L3 type acute lymphoblastic leukemias, and plasmacytoma (141,187,301). H-DNA is an intra-molecular triplex structure formed at alternating purine-pyrimidine mirror repeats (Fig. 3.1) (21). We found that a 23-bp H-DNA-forming sequence from a Burkitt's lymphoma translocation hotspot in *c-MYC* (21) stimulated DNA double-strand break (DSB) formation and induced mutations, deletions, and translocations in mammalian cells and mice (125,128). However, the mechanisms involved in H-DNA-induced genetic instability are largely unknown.

The formation of the H-DNA structure causes significant helical distortion, which may mimic a bulky DNA adduct, similar to substrates of several DNA repair pathways, such as nucleotide excision repair (NER) and mismatch repair (MMR). Thus, we *hypothesize* that proteins involved in NER, MMR and structure-specific DNA processing are likely involved in H-DNA-induced mutagenesis.

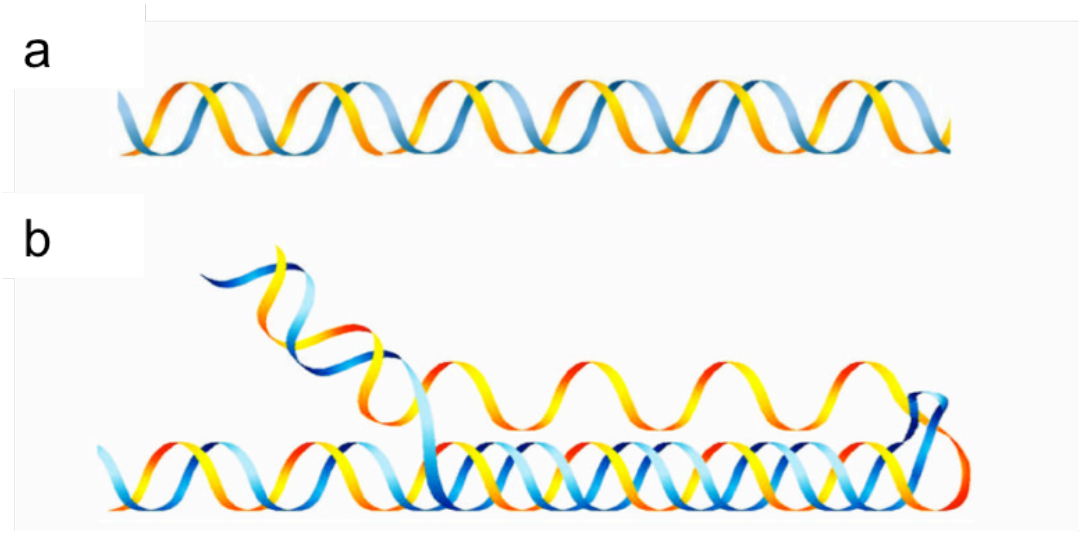


Figure 3.1. DNA structures. (a) Watson-Crick double-stranded B-DNA. **(b)** H-DNA intra-molecular triplex structure [adapted from (256)].

To evaluate the involvement of repair proteins in H-DNA-induced genetic instability, we developed a yeast artificial chromosome (YAC)-based genetic screen and found that Rad1(XPF) was required for H-DNA-induced mutagenesis. Additionally, we confirmed these results in human cells using a plasmid-based assay, and performing *in vitro* assays, and found that XPF was involved in the formation of DSBs, and had the capacity to cleave H-DNA structures along with both XPG and FEN1. Interestingly, FEN1 was found to be critical for maintaining genome stability during replication.

Rad1(XPF) is a structure-specific nuclease required for NER, an important pathway for repairing a wide variety of structurally diverse DNA lesions. Once DNA damage is recognized and verified by the NER complexes XPC-RAD23B and XPA-RPA, the ERCC1-XPF complex incises the double-stranded DNA (dsDNA) 5' to the lesion, while XPG incises the dsDNA 3' to the damage. Both ERCC1-XPF and XPG are structure-specific endonucleases with substrate preferences such as NER pre-incision bubbles containing dsDNA/ssDNA junctions, stem loops, and 3' and 5' flaps respectively (302-304).

Naturally occurring H-DNA (i.e. *intra*-molecular triplex) structures cause distortions to the DNA backbone, raising the possibility that H-DNA may be recognized as "damaged DNA" by the NER machinery. In support of this notion, NER damage/distortion recognition complexes XPA-RPA and XPC-RAD23B bind inter-molecular DNA triplexes, which are structurally similar to intra-molecular H-DNA triplexes (31), with high affinity (120,126). Here, we identified H-DNA as a

novel substrate for the NER endonucleases ERCC1-XPF and XPG independent of DNA replication, and that these nucleases are involved in H-DNA-induced genetic instability, in the presence or absence of DNA replication. In contrast, we identified a novel function for another XPG-family endonuclease, FEN1, in *preventing* H-DNA-induced genetic instability in a *replication-dependent* manner. During replication, FEN1 is responsible for removing the 5' flaps of Okazaki fragments and thereby prevents chromosome breaks and deletions (261,305,306). In support of this genome maintenance function, mutations in FEN1 can lead to genetic instability (307) and homozygous knockout of *FEN1* is embryonic lethal in mice (308), underscoring the importance of FEN1 in proliferating cells. Our results of replication-independent genomic instability induced by H-DNA in FEN1 knockdown cells, taken together with the role of ERCC1-XPF and XPG in the presence or absence of replication provides, for the first time, distinct nuclease-associated, replication-dependent and replication-independent mechanisms of DNA structure-induced genetic instability in eukaryotes.

3.3. Materials and Methods

3.3.1. Bioinformatics analyses. The dataset containing the breakpoint locations of cancer genomes translocations and deletions mapped to the human genome assembly GRCh37/hg19 was obtained from COSMIC at http://cancer.sanger.ac.uk/. The total number of unique breakpoints for translocations and deletions was 19957 and 46372, respectively. Triplex-forming repeats (TFRs) located within ±100 bp from the breakpoints (bins), comprising two adjacent purines or pyrimidine runs of length ≥6 bases each, displaying mirror symmetry, and separated by a loop size of 0-7 bases, were identified using custom scripts. For overlapping TFRs only the longest match was output. The bedtools utility was used to generate a set of 20,000 non-gap-matching sequences (bins), each 200 bases long (166).

3.3.2. Yeast strains and YACs. Yeast strain 213 (MATa *Kar1-1, his7, leu2-3, 112, ura3-52*) was used for YAC construction and transfer into other yeast strains. The yeast strain BY4742 (*MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*) and derivatives (Yeast deletion library GSA-5, ATCC) were used in mutation screening. H-DNA or control sequences were cloned between the telomere seed G4T4 and the *URA3* gene on a replication-defective plasmid derived from pRS306. Plasmids were linearized and used to construct YACs by homologous recombination with YAC VS5 containing a point mutation in the *URA3* gene

(261,263). H-DNA or control sequences and a functional *URA3* gene were selected on yeast minimal media SD base (Clontech, Mountain View, CA) with CSM-Ura (MP Biomedicals, Santa Ana, CA) plates and confirmed by polymerase chain reaction (PCR) with primers T7 and T3 (Table 3.1), followed by direct DNA sequencing.

3.3.3. YAC transfer (kar-cross). Donor cells K213 containing YACs with human H-DNA or control sequences were grown from a single colony in SD base with CSM-Ura-Leu medium (MP Biomedicals) overnight at 30°C. Canavanine resistant recipient cells from the yeast deletion library (BY4742 background) were grown in yeast complete media YPD at the permissive temperature overnight. YACs were transferred from donor cells to recipient cells via *kar*-mediated transfer as described in Chapter 2 (Fig. 2.5) and previously (260,261). *Kar* crossed colonies were replica-plated onto CSM-Ura-Leu and CSM-lys plates to confirm YAC transfer (i.e.: colonies positive for YAC would grow on CSM-Ura-Leu, but not on CSM-Lys plates). Insertion sequences on YACs were further confirmed by PCR amplification with primers T7 and T3 (Table 3.1) followed by sequencing.

3.3.4. YAC fragility assay. Single yeast cells harboring YACs containing human H-DNA or control sequences were used to inoculate cultures and were grown for 20 hours at the permissive temperature in SD base with CSM-Leu

medium. 50 µL of each culture was plated on the SD base with CSM-Leu plates with 5-fluoroorotic acid (5-FOA, Zymo Research, Irvine, CA) to select for breakage events. 10 µL of each sample was diluted and plated on the SD base with CSM-Leu plates for a total cell number count. The mutation frequencies were calculated as the number of FOA resistant (FOA^R) colonies divided by the number of total colonies (Fig. 2.6).

3.3.5. PCR and Southern blot analysis of FOA^R YAC structure. FOA^R colonies were harvested in 30 µL water, boiled at 95°C for 5 min to lyse the cells, and 6 µL were used as templates for PCR reactions with URA3For and URA3Rev, and T3 and T7 primer sets (Table 3.1). Amplification products were separated on a 1% agarose gel to detect the mutations/deletions in the *URA3* gene in the YAC. Samples positive for the PCR product (i.e. the *URA3* region) were gel purified using a Qiagen gel purification kit, and analyzed by direct DNA sequencing to confirm *URA3* mutation resulting in FOA^R. Southern blotting was performed on FOA^R colonies as previously described (261,263). Briefly, FOA^R colonies were grown 16 to 24 hours in CSM-Leu media, and genomic DNA was isolated using glass beads and Zymolyase (Zymo Research, Irvine, CA). Genomic DNA was digested with *Bst*EII (New England Biolabs), separated on a 0.8% agarose gel at 45 volts overnight, probed with a digoxigenin-labeled (Roche) probe of lambda DNA digested with *Hind*III (NEB), and detected by a chemiluminescent system (Roche). To detect the H-DNA sequences in the right

arm of YAC, genomic DNA samples were digested by *Pvu*II (NEB) and separated on a 1.0% agarose gel at 45 volts overnight. A digoxigenin-labeled specific H-DNA probe was used, and the blot was detected using a chemiluminescent system as described (Callahan MCB 2003) (Roche).

3.3.6. H-DNA-induced mutagenesis assay in human cells. HeLa cells, human XPF-proficient (GM08437B-XPF) or XPF-deficient cells (GM08437B-pLPC) (266,267), and human XPG-proficient or XPG-deficient cells (309) were maintained in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, CA) with 10% fetal bovine serum (FBS) and antibiotics. Human XPA-proficient (XP12RO) and XPA-deficient (XP12RO Clone12) cells (310) were maintained in RPMI medium with 10% FBS and antibiotics. Human MSH2-deficient (HEC59) and isogenic control MSH2-proficient (HEC59 + Chr2) cells [provided by Dr. Randy J. Legerski, The University of Texas, M.D. Anderson Cancer Center (311,312)] were maintained in DMEM with 10% FBS and antibiotics; MSH2-proficient selection was maintained in DMEM with 10% FBS and 600 µg/mL G418•Sulfate (Enzo Life Sciences, Farmingdale, NY). Human MLH1-deficient (A2780-MNUCL1) and isogenic control MLH1-proficient (MLH1-1) cells [provided by Dr. Margherita Bignami, Instituto Superiore di Sanita, Rome, Italy (313-315)] were maintained in DMEM with 10% FBS and 500 µg/mL G418•Sulfate. The H-DNA-containing (pMexY) or control plasmids (pCex) were transfected into human cells using GenePORTER (Genlantis Inc., San Diego,

CA) according to the manufacturer's recommendations. *E. coli* MBM7070 cells (F-*lacZ (am)CA7020, lacY1, hsdR−, hsdM+, araD139 Δ(araABC-leu)7679, galU, galK, rpsL, thi*) (316) were used in the *supF*-based mutation frequency assays. Shuttle vectors carrying human H-DNA or control B-DNA sequences were extracted from human cells and transformed into MBM7070 cells mutation frequencies were assayed by blue/white screening as previously described (262). DNA from individual mutant colonies was sequenced to characterize the mutations/deletions near the H-DNA or control B-DNA regions.

3.3.7. siRNA knockdown in cultured human cells. Human FEN1 or non-targeting siRNA (Dharmacon, GE Healthcare) were transfected into cultured XPF-proficient or XPF-deficient human cells with RNAimax (Invitrogen, Life Technologies) using the manufacturer's recommended protocol. A second siRNA transfection together with the mutation-reporter plasmids was carried out 48 hours after the first transfection using GenePORTER (Genlantis). Cells were harvested 48 hours later for Western blotting to examine the FEN1 protein levels, and plasmids were extracted for mutation analysis.

3.3.8. Replication effect on H-DNA-induced mutation frequency. The SV40 origin of replication in the plasmids carrying the human H-DNA-forming or control B-DNA sequences was deleted by restriction digestion (using *Nco*1) and ligation. Plasmids with or without the SV40 replication origin (pMexY and pCex; pMexY-

SV40 and pCex-SV40, respectively) were transfected into human cells using a Nucleofector (Lonza, Walkersville, MD). Cells were collected 24 or 48 hours after transfection for blue/white screening in *E. coli* MBM7070 cells as described above.

3.3.9. Ligation mediated-PCR (LM-PCR) analysis of H-DNA-induced DSBs. Plasmids were recovered from human cells using a modified Hirt extraction method that we have described previously (125,262). Briefly, extracted plasmids were treated with Pol I Klenow fragment to blunt the DNA ends, and ligated to the PCR Linkers (annealed from oligonucleotides LMPCR1 and LMPCR2). The ligation products were used as templates in the PCR reaction with a specific primer (LMsupF183) to the *supF*-reporter gene and a linker primer (LMPCR2) (Table 3.1). Amplified products were separated on a 2% agarose gel, and fragments of interest were purified from the gel and sequenced to map the DSB sites.

3.3.10. Chromatin immunoprecipitation (ChIP) assay. ChIP assays were performed using a Simple Chip Enzymatic Chromatin IP Kit (Cell Signaling, Inc., Santa Cruz, CA) according to the manufacturer's recommended protocol and as described previously (130). Some modifications in protocols were specific to yeast cells. Yeast cells containing YACs with human H-DNA-forming or control B-DNA sequences were grown overnight at a permissive temperature and 5 mL of

culture was fixed with 1% formaldehyde for 8 min at room temperature to crosslink DNA and proteins. After quenching and washing, the YACs from the yeast cells were extracted using lysis buffer (0.1 M Tris-Cl pH 8.0, 50 mM EDTA, 1% SDS) and Zymolyase (Zymo Research), followed by sonication to obtain an average DNA fragment length <500 bp. Approximately 1% of the total chromatin was stored as "input DNA" and the remaining chromatin was incubated with 5 μg of specific antibody (α-FEN1, α-XPG, α-XPF, Abcam, Cambridge, MA) or control α-IgG antibody and samples were incubated overnight at 4°C on a rotator. The rest of steps were performed according to the ChIP Kit protocols. Fractions of purified ChIP DNA and input DNA were used for PCR analysis. The primers for PCR amplification were T3 and T7 surrounding the H-DNA-forming or control sequences (Table 3.1). Amplified products were separated on 1.5% agarose gels containing ethidium bromide and visualized on a ChemiDoc (Bio-Rad Laboratories, Hercules, CA).

Table 3.1. List of oligonucleotides used in this study.

Name	Sequence
MCR2-5'	5'-cag gaa atc acc cct ccc ttt ttg gga ggg gcg ctt atg ggg agg g
71	5'-ggt acc gaa ttt cgg ccg agg ggg agg ggg tgg tgg ggg ggg aag gat tcg aac ctt
72	5'-aag gtt cga atc ctt ccc ccc cca cca ccc cct ccc cct cgg ccg aaa ttc ggt acc
FENS1	5'-gtc atg ata gat ctg atc gct cga att cct gca gcc cgg
FENS2	5'-ccg ggc tgc agg aat tcg ata tca agc tta tcg ata ccg tcg acc tcg a
FENS3	5'-tcg agg tcg acg gta tcg ata agc ttg ata
T7	5'-taa tac gac tca cta tag gg
T3	5'-att aac cct cac taa agg ga
URA3For	5'-tgc tgc tac tca tcc tag
URA3Rev	5'-tcc cag cct gct ttt ctg ta
LMPCR1	5'-cgt aca ttc aca acg ata gcg act ga
LMPCR2	5'-gct atc gtt gtg aat gta cg
LMsupF183	5'-aga tcc agt tcg atg taa cc

[modified from Zhao, *et al.* (submitted *Nature Communications* 2016)]

3.3.11. Substrate preparation. The H-DNA substrate was formed by a single stranded oligonucleotide MCR2-5' (Table 3.1); the duplex DNA substrate was formed by annealing oligonucleotides 71 and 72 (130); the FEN1 cleavage control substrates were formed by annealing oligonucleotides FENS1, FENS2, and FENS3 (317); the stem-loop substrate was formed by a single-stranded oligonucleotide SL46 (302). The H-DNA fold-back oligonucleotide MCR2-5' was incubated in triplex-forming buffer (20 mM NaCacodylate, 0.1 mM EDTA, 10 mM $MgCl_2$, 100 mM NaCl), and the other substrates were formed by annealing oligonucleotides in buffer (10 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA), boiled for 10 min, and slowly cooled to room temperature. Formation of the various structures was confirmed by native acrylamide gel electrophoresis and/or circular dichroism.

3.3.12. Circular dichroism (CD). Blank and sample-containing quartz cuvettes were matched prior to use. The CD spectrum of 10 μM MCR2-5' was measured at pH 7.0 in 20 mM HEPES, 100 mM NaCl, 0.1 mM EDTA and 10 mM $MgCl_2$ at 25°C. Samples were evaluated on a Jasco J-815 CD Spectrometer (JASCO, Inc., Easton, MD) equipped with a Peltier temperature controller and scanned from 200 – 350 nm. Data were processed using the JASCO software and exported to Sigma Plot 10. The negative peak at 210-220 nm represents the formation of a triplex structure.

3.3.13. Cleavage assays. Radiolabeled H-DNA-forming fold-back oligonucleotides, duplex DNA, and flap substrates (at $6x10^{-8}$ M) were incubated with 20 ng of purified human recombinant FEN1 protein (Provided by Dr. Binghui Shen, City of Hope) in reaction buffer (40 mM Tris-HCL pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 200 mg/mL BSA) (317) at 30°C for 15 min. Cleaved products were separated on a 20% denaturing polyacrylamide gel, and visualized using a Typhoon PhosphorImager (GE Healthcare Life Sciences). Forty-eight ng of purified human recombinant XPG protein (provided by Dr. Richard D. Wood, The University of Texas M.D. Anderson Cancer Center) (318), or BSA was incubated with $4x10^{8}$ M DNA substrates at 30°C overnight in a buffer containing 25 mM HEPES (pH=6.8), 4 mM $MnCl_2$, 1 mM DTT, 250 µg/mL BSA, and 10% glycerol. Formamide/dye was added to stop the reactions and the cleavage products were resolved on a 12% denaturing polyacrylamide gel. Cell extracts were prepared from XPF-proficient or XPF-deficient cells (266,267) using a NucBuster kit (EMD Millipore, Temecula, CA). The triplex structure formed by oligonucleotide MCR2-5' (Table 3.1) was incubated with cell extract in reaction buffer (40 mM Tris-HCL pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 200 mg/mL BSA) at 30°C for 1 hour. Formamide/dye was added to stop the reactions and the cleavage products were resolved on a 12% denaturing polyacrylamide gel.

3.4. Results

3.4.1. H-DNA-forming sequences map to breakpoints in human cancer genomes.

To determine the impact of H-DNA-forming sequences in human cancer etiology, we mapped H-DNA (triplex)-forming sequences (TFRs) within ±100 bp (thereafter referred as to bins) of 19,956 translocation breakpoints from sequenced cancer genomes (COSMIC at http://cancer.sanger.ac.uk/). We determined the distributions of each TFR midpoint relative to the breakpoint positions (taken to be 0). Both the fraction of bins harboring TFRs ≥6 bp on each mirror repeat arm and the number of TFRs (normalized to 20,000 bins) were greater for translocations (35.3±6.6 and 6,676) than for those in 20,000 random-pick control sequences extracted from the reference human genome (24.9±4.9 and 4,715) (*P* value from t-test, 4.2×10^{-50}, Fig. 3.2a). In addition, the number of TFRs in each bin surrounding the translocation breakpoints was consistently higher than control (Fig. 3.2b). Thus, H-DNA-forming sequences are significantly enriched surrounding human cancer translocation breakpoints, implicating H-DNA in chromosomal instability and cancer development (166).

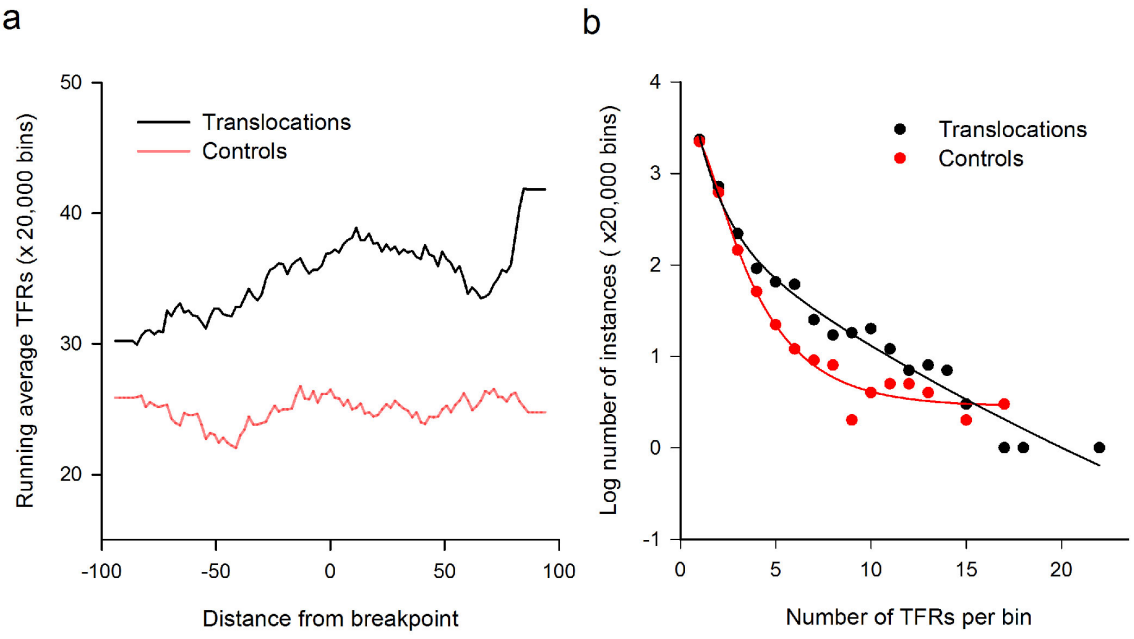


Figure 3.2. H-DNA (triplex)-forming sequences (TFRs) are associated with translocations in human cancer genomes. (a) Running average of TFRs in the human cancer translocation breakpoint (red) and control (black) datasets whose center loop positions are located at each base along the ±100 bp flanking cancer translocation breakpoints or the control sites. **(b)** Distribution of number of bins/20,000 bins containing TFRs in cancer translocation breakpoint (red) and control (black) datasets [adapted from (166)].

3.4.2. H-DNA-induced genetic instability in Saccharomyces cerevisiae.

To identify the gene products involved in H-DNA-structure-induced DSB formation and genetic instability, we developed a yeast artificial chromosome (YAC) (261,263) carrying a 23-bp human H-DNA-forming sequence from a Burkitt's lymphoma translocation hotspot in *c-MYC* or a B-DNA control sequence (Fig. 3.3a) adjacent to a selectable *URA3* gene (266). We found for the first time, that this human H-DNA-forming sequence stimulated genetic instability ~9.4-fold above that of the control in wild-type (WT) BY4742 yeast (Fig. 3.3b, 11.60×10^{-5} vs. 1.24×10^{-5}). PCR (Fig. 3.3c) and Southern blotting (Fig. 3.3d) confirmed that the YAC underwent DSBs and lost the distal end of the chromosome in the majority of the URA3-deficient cells, consistent with our finding of H-DNA-induced DSBs and genetic instability in mammalian cells and mice (125,128). Thus, the H-DNA-induced mutagenesis in yeast is comparable to that detected in human cells, and likely occurs via a similar processing mechanism.

We used this YAC system to screen multiple repair-deficient strains for effects on H-DNA-induced mutagenesis when compared to the ~9.4-fold induction seen in WT cells. Due to the similarity of structural characteristics of H-DNA to substrates processed by the NER and MMR mechanisms (i.e. bulky adducts, and helical distortions), cells deficient in these repair pathways were initially screened. Data obtained for the NER pathway are discussed below as included in our manuscript. Please see the section titled, "Work in progress

related to the H-DNA-induced mutagenesis project" at the end of this chapter for the MMR-related data.

Strikingly, in yeast *rad1Δ(XPF)* and *rad10Δ(ERCC1)* strains that are deficient in the NER endonuclease complex Rad1-Rad10 (XPF-ERCC1), H-DNA-induced genetic instability was significantly reduced to 37% and 24% of that found in the WT BY4742 strain, respectively (Fig. 3.3b; *t*-test, *P* value <0.05). Thus, in *rad1Δ and rad10Δ* cells H-DNA-induced genetic instability was only 3.3-fold and 2.0-fold above that of the control B-DNA sequence, significantly lower than the 9.4-fold induction in the WT BY4742 cells, indicating that the NER nuclease complex, Rad1-Rad10(ERCC1-XPF), is necessary for the H-DNA-induced chromosomal breaks and genetic instability in yeast.

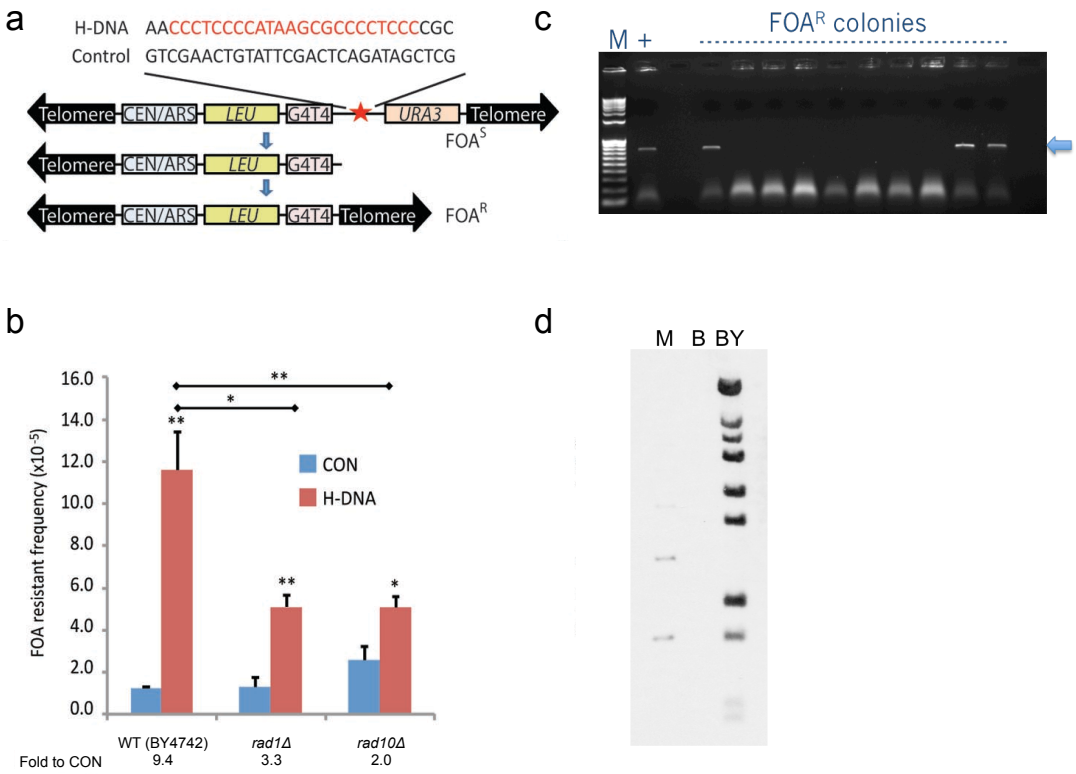


Figure 3.3. NER-associated H-DNA-induced genetic instability in yeast cells. (a) Schematic structure of YACs used for H-DNA-induced genetic instability in yeast. **(b)** H-DNA-induced fragility frequencies in reporting YACs, as assessed by 5-flurooratic acid (FOA) resistance, in WT (BY4742), *rad1Δ* or *rad10Δ* yeast strains from a yeast single-gene deletion library (GSA-5, ATCC). The numbers below the graph show the fold increases in H-DNA-induced genetic instability above the control in each cell line (Fold to Con). At least 30,000 colonies were screened in each fragility assay. Error bars represent the s.d. from three repeats; two-sided *t*-test, *P* values <0.05(*), <0.01(**). **(c)** Representative agarose gel for analysis of FOA^R colonies that were lysed and used as PCR templates using URA3For and URA3Rev primer sets. Products were separated on a 1% agarose gel with a 1 kb and 100 bp marker (M). Starting material for the fragility assay containing full length, FOA^S YAC, served as a positive control (+). The blue arrow indicates the target region amplified. **(d)** Representative Southern blot analysis to determine presence of YAC in FOA^R colonies. Isolated genomic DNA from colonies containing the YAC (BY) were digested with *Bst*EII, separated on a 0.8% agarose gel with a 1-kb ladder (M), and probed with a digoxigenin-labeled probe of lambda DNA digested with *Hind*III and detected by chemiluminescence. Cells lacking the YAC served as a negative control (B). These experiments were performed by Jennifer McKinney and Dr. Junhua Zhao [Zhao, *et al.* (submitted *Nature Communications* 2016)].

3.4.3. H-DNA-induced genetic instability in human cells.

To determine the effect(s) of NER factors on H-DNA-induced mutagenesis in human cells, we utilized a *supF*-reporter system (125) (Fig. 3.4a) in human cells from Xeroderma Pigmentosum (XP) patients (318). Consistent with our results in yeast, H-DNA-induced genetic instability was significantly decreased in XPF-deficient human cells to that of isogenic XPF-proficient (WT) cells (2.0-fold induction over the control compared to 11.1-fold in the human XPF-proficient cells) (Fig. 3.4b). A similar reduction in H-DNA-induced mutagenesis was also observed in XPA-deficient patient cells, resulting in a decrease of H-DNA-induced mutation frequency from 11.1-fold in XPA-proficient (WT) cells over the control to a 2.9-fold induction over control in XPA-deficient cells (Fig. 3.4b), suggesting a role for functional NER in the mutagenic processing of H-DNA in human cells.

Analysis of the H-DNA-induced mutation spectra in human cells revealed that XPF-deficiency substantially reduced deletion events from 77% (28/36 mutants sequenced) in XPF-proficient cells to 41% (12/29 mutants sequenced) in XPF-deficient cells (Table 3.2; x^2 test, *P* value <0.01), suggesting that XPF is involved in H-DNA processing in human cells. In addition to ERCC1-XPF, XPG is a structure-specific endonuclease required for cleaving damaged DNA in NER. Although XPG deficiency did not substantially alter the H-DNA-induced mutation frequency (data not shown), its absence resulted in significantly fewer deletion

events compared to XPG-proficient human cells; 71% (15/21) of the mutants contained deletions in the XPG-proficient cells versus 33% (7/21) in the XPG-deficient cells (Table 3.2; x^2 test, *P* value <0.05). Collectively, these results implicated the NER mechanism in H-DNA-induced genetic instability in yeast and human cells.

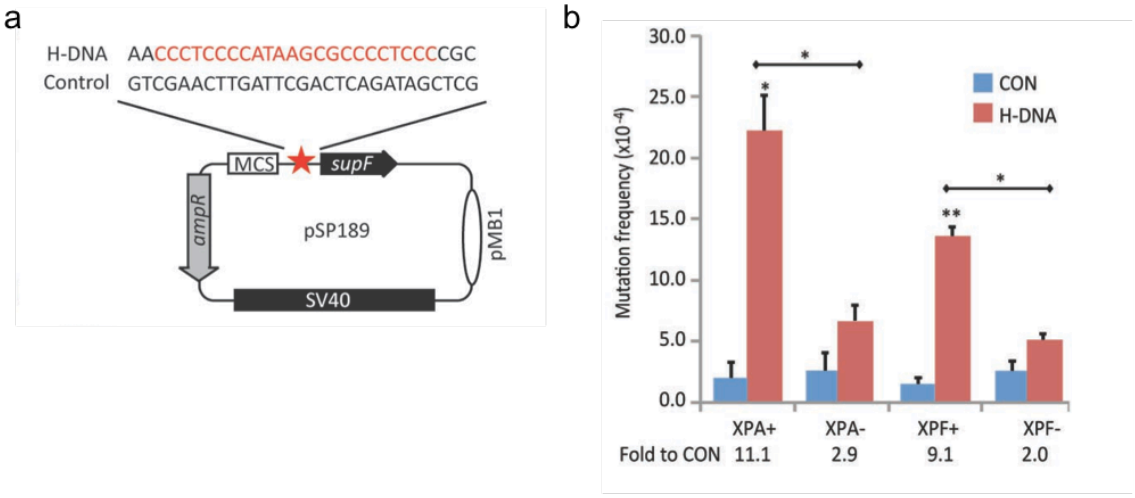


Figure 3.4. NER-associated H-DNA-induced genetic instability in human cells. (a) Schematic structure of the *supF*-reporter vector containing human B-DNA or H-DNA-forming sequences. **(b)** H-DNA-induced mutation frequencies in NER-proficient (WT) and NER-deficient human cells. "Fold to Con" was calculated as described above in Figure 3.2b. At least 30,000 colonies were screened in each assay. Error bars represent the s.d. from three repeats; two-sided *t*-test, *P* values <0.05(*), <0.01(**). These experiments were performed by Jennifer McKinney and Dr. Junhua Zhao [Zhao, *et al.* (submitted *Nature Communications* 2016)].

Table 3.2. Analysis of H-DNA-induced and spontaneous (B-DNA) mutations in human cells.

Strains	Replication status	H-DNA		B-DNA control	
		Deletions	Point mutations	Deletions	Point mutations
XPF+	+	77.8% (28 /36)	22.2% (8/36)	38.9% (7/18)	61.1% (11/18)
XPF-	+	41.4%** (12/29)	58.6% (17/29)	53.8% (14/26)	46.2% (12/26)
XPG+	+	71.4% (15/21)	28.6% (6/21)	41.2% (7/17)	58.8% (10/17)
XPG-	+	33.3%* (7/21)	66.7% (14/21)	12.5% (2/16)	87.5% (14/16)
XPA+	+	71.4% (10/14)	28.6% (4/14)	11.1% (2/18)	88.9% (16/18)
XPA-	+	56.5% (13/23)	43.5% (10/23)	26.7% (4/15)	73.3% (11/15)
HeLa, siCON	+	77.8% (14/18)	22.2% (4/18)	66.7% (6/9)	33.3% (3/9)
HeLa, siFEN	+	100.0% **(43/43)	0.0% (0/43)	92.9% (13/14)	7.1% (1/14)
XPF+, siCON	-	15.0% (3/20)	85.0% (17/20)	0.0% (0/21)	100.0% (21/21)
XPF+, siCON	+	47.4% (9/19)	52.6% (10/19)	4.4% (1/23)	95.6% (22/23)
XPF-, siCON	-	12.5% (1/8)	87.5% (7/8)	19.0% (4/21)	81.0% (17/21)
XPF-, siCON	+	34.8% (8/23)	65.2% (15/23)	0.0% (0/19)	100.0% (19/19)
XPF+, siFEN	-	30.8% (4/13)	69.2% (9/13)	25.0% (2/18)	75.0% (16/18)
XPF+, siFEN	+	28.0% (7/25)	72.0% (18/25)	6.2% (1/16)	93.8% (17/16)
XPF-, siFEN	-	14.3% (1/7)	85.7% (6/7)	83.3% (10/12)	16.7% (2/12)
XPF-, siFEN	+	15.8% (3/19)	84.2% (16/19)	5.3% (1/19)	94.7% (18/19)

"*": *P* value <0.05; "**": *P* value <0.01 in χ^2 test.
[Zhao, *et al.* (submitted *Nature Communications* 2016)]

3.4.4. H-DNA is a substrate for NER-associated endonucleases.

Because we detected a significant decrease in H-DNA-induced deletions in the absence of ERCC1-XPF (Table 3.2), we speculated that this NER structure-specific nuclease was involved in cleaving the H-DNA structure during its processing. Thus, to determine the effect of ERCC1-XPF on H-DNA-induced DSB formation, ligation-mediated PCR (LM-PCR) was employed to map DSBs in human cells. Our LM-PCR results identified DSBs near the H-DNA-containing region on plasmids recovered from human XPF-proficient cells with a particularly strong signal at 230 bp (Fig. 3.5a) near the H-DNA site. In contrast, the DSB signal at the H-DNA site (230 bp) was substantially diminished in the XPF-deficient human cells (Fig. 3.5a, compare lane 4 to lane 3), further supporting a role for ERCC1-XPF in processing H-DNA structures to generate DSB intermediates, resulting in deletions and subsequent genetic instability.

Preferred substrates for ERCC1-XPF and XPG cleavage include NER pre-incision bubbles containing dsDNA/ssDNA junctions, stem loops, and 3' and 5' flaps (302-304); however, the activities of ERCC1-XPF and XPG on H-DNA are unknown. In the XPF-proficient human cell extracts, an H-DNA structure formed on oligonucleotide MCR2-5' (Table 3.1 and Fig. 3.5b) (confirmed by native gel electrophoresis and circular dichroism, Fig. 3.6) was cleaved resulting in a 37-nt cleavage product, which was dramatically reduced in the XPF-deficient extracts (Fig. 3.5c). Interestingly, the predominant cleavage site was on the loop between

the two Hoogsteen hydrogen-bonded strands, rather than at the Watson-Crick hydrogen-bonded region adjacent to the loop. Consistent with a role for NER in H-DNA processing, purified human recombinant XPG also cleaved the MCR2-5' H-DNA substrate, resulting in removal of the 5' ssDNA flap at the junction adjacent to the Watson-Crick base-paired duplex (Fig. 3.5d). To our knowledge, this is the first demonstration of cleavage of H-DNA by ERCC1-XPF and XPG, and thus represents a novel substrate for these nucleases, further implicating NER in the mutagenic processing of H-DNA.

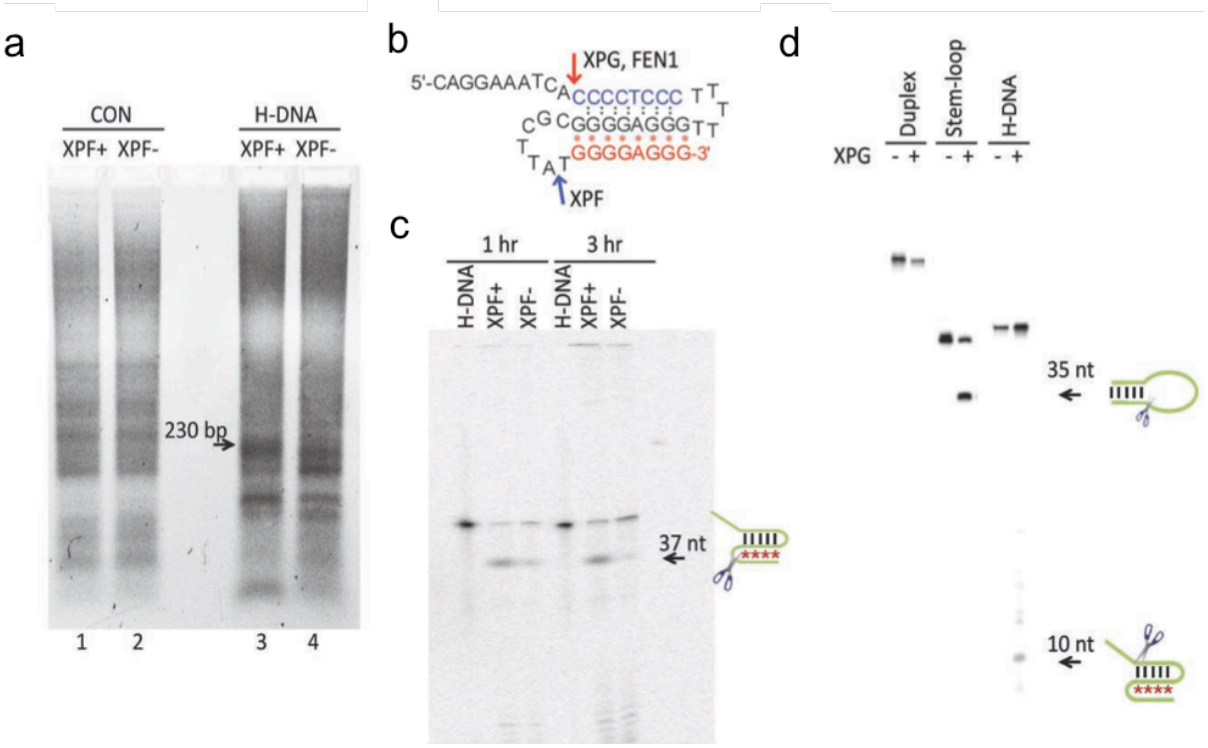


Figure 3.5. NER-associated H-DNA-induced processing. (a) LM-PCR detection of DSBs on plasmids recovered from XPF-proficient or XPF-deficient human cells 24 hours after transfection. PCR products were separated on a 1.5% agarose gels. The 230-bp PCR fragment maps to a DSB hotspot within the H-DNA-forming sequence. **(b)** Schematic of the H-DNA structure of oligonucleotide MCR2-5'. Black dots represent the Watson-Crick hydrogen-bonds, and the red * represent the Hoogsteen hydrogen bonds at the triplex structure. Cleavage sites of XPF, XPG, and FEN1 from (c), (d) and Fig. 3.7c are indicated on the structure in the schematic. **(c)** Cleavage of H-DNA in XPF-proficient and XPF-deficient human cell extracts. The 37-nt cleavage product is marked by an arrow. **(d)** H-DNA cleavage by purified recombinant human XPG protein. A stem-loop substrate serves as a positive control for XPG cleavage. The cleavage products of 35-nt from the stem-loop, and 10-nt from the H-DNA substrates are marked by arrows. These experiments were performed by Dr. Junhua Zhao. [Zhao, *et al.* (submitted *Nature Communications* 2016)].

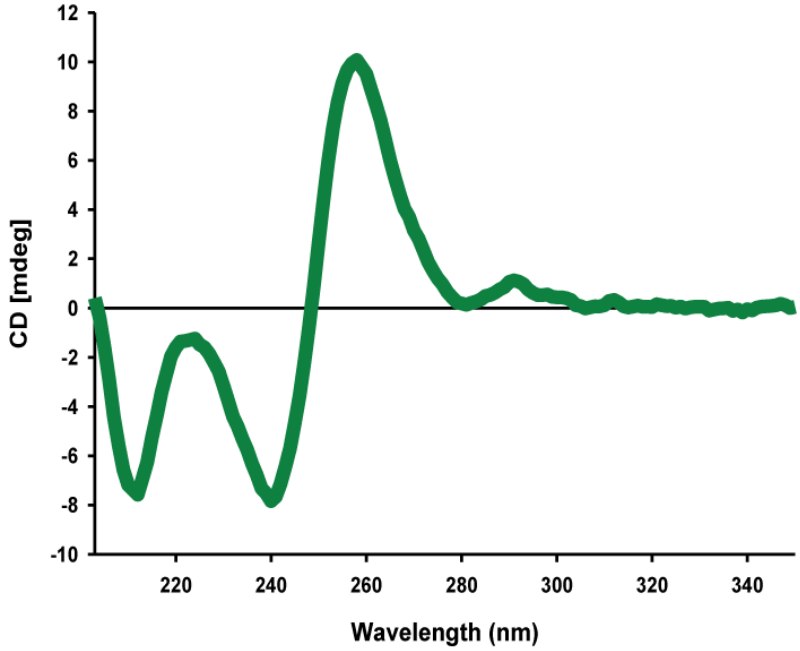


Figure 3.6. Circular dichroism (CD) spectrum of the triplex fold-back oligonucleotide MCR2-5'. The CD spectrum of MCR2-5' at a 10 μM concentration in a buffer (pH 7.0) containing 20 mM HEPES, 100 mM NaCl, 0.1 mM EDTA and 10 mM $MgCl_2$, at 25°C. The strong negative peak at ~210 nm is characteristic of triplex formation (319). This experiment performed by Dr. Imee Del Mundo. [Zhao, *et al.* (submitted *Nature Communications* 2016)].

3.4.5. Rad27(FEN1) inhibits genomic instability caused by H-DNA.

Based on our results with the NER nucleases, we speculated that other structure-specific endonucleases may also process H-DNA structures. For example, FEN1 belongs to the FEN1/XPG family of endonucleases, and shares a conserved active site with XPG (320). Thus, we evaluated H-DNA-induced mutagenesis in a *rad27*-deleted yeast strain, the yeast homolog of human *FEN1*. Strikingly, H-DNA-induced mutagenesis was substantially *increased* in the *rad27Δ* strain relative to the WT cells (Fig. 3.7a: 11.6×10^{-5} in WT versus 56.5×10^{-5} in the *rad27Δ* cells; *t*-test, *P* value <0.01), ~794% over that induced by H-DNA in the WT strain (*t*-test, *P* value <0.01). Consistent with our results in yeast, depletion of FEN1 using siRNA in HeLa cells (to <10% of WT levels as assessed by immunoblotting, Fig. 3.8) increased H-DNA-induced genetic instability over the control from 6.5-fold in WT HeLa cells to 19.7-fold in the FEN1-depleted HeLa cells, ~485% over that induced by H-DNA in the WT cells (Fig. 3.7b; *t*-test, *P* value <0.05). These results implicate FEN1 in maintaining genomic stability at the H-DNA region, perhaps by cleaving and unwinding the mutagenic H-DNA structure.

Further support for this notion was provided by our finding that human recombinant FEN1 protein (321) cleaved the MCR2-5'-derived H-DNA substrate at the base of the 5' flap, resulting in a 10 nt cleavage product (Fig. 3.7c). Moreover, FEN1-depletion in HeLa cells resulted in a statistically significant

increase in large deletions (100%, 43/43) compared to the WT cells (Table 3.2; x^2 test, *P* value <0.01). Thus, we have identified H-DNA as a novel substrate for FEN1 cleavage, and in contrast to cleavage by the NER nucleases, FEN1 cleavage resulted in reduced DNA structure-induced mutations.

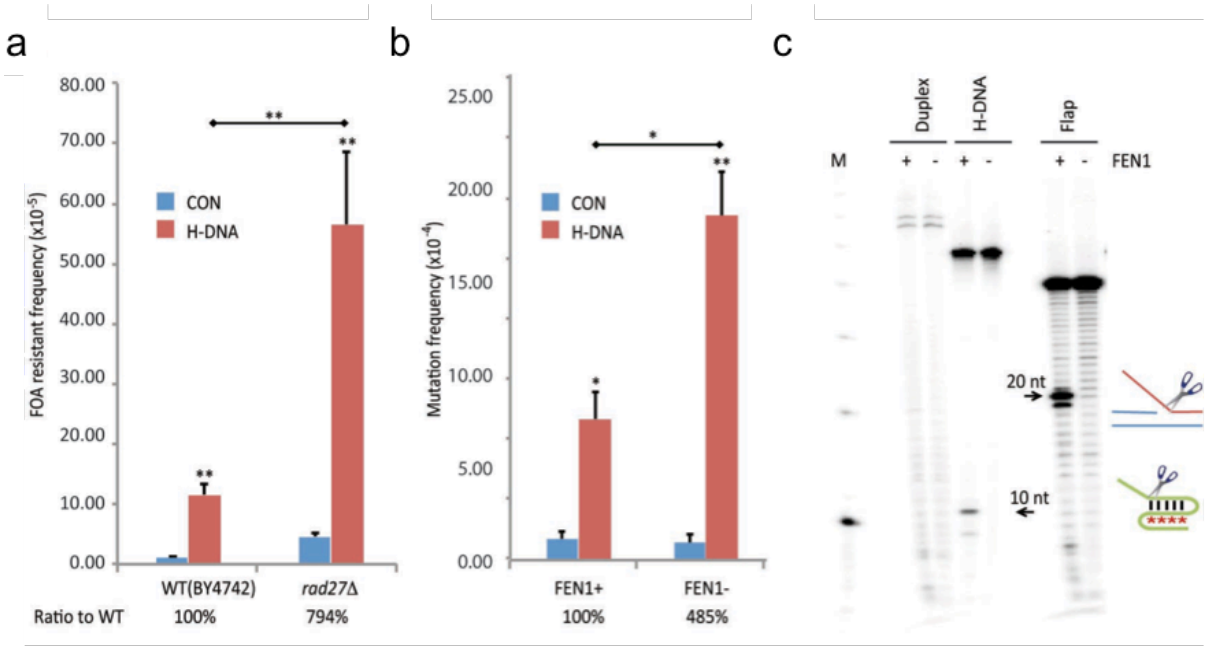


Figure 3.7. FEN1 cleavage and inhibition of H-DNA structure-induced mutagenesis. (a) H-DNA-induced FOA resistance on YACs in WT (BY4247) or *rad27Δ* strains. **(b)** H-DNA-induced mutation frequencies on *supF*-mutation reporters in WT or FEN1-depleted HeLa cells. **(c)** Cleavage of H-DNA formed on oligonucleotide MCR2-5' and flap structures (Table 3.1) by purified recombinant human FEN1 protein. The 20-nt cleavage product from its preferred substrate (a 5' flap), and the 10-nt product from H-DNA are marked by arrows. These experiments were performed by Jennifer McKinney and Dr. Junhua Zhao. [Zhao, *et al.* (submitted *Nature Communications* 2016)].

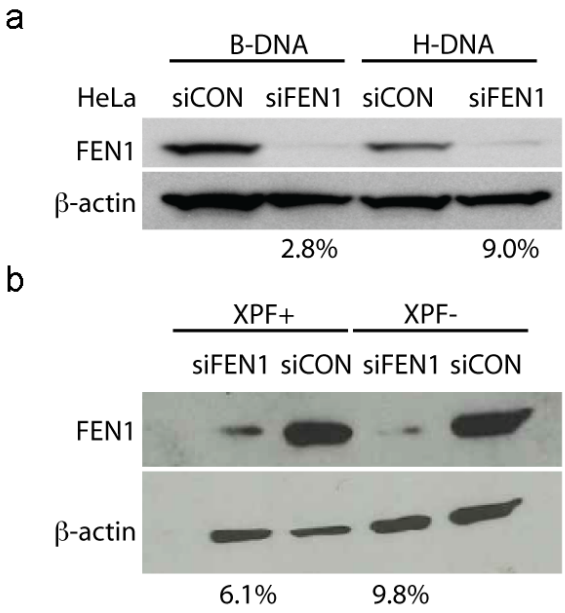


Figure 3.8. SiRNA-mediated depletion of FEN1 in human cells. (a) Detection of FEN1 protein in HeLa cells treated with FEN1 siRNA (siFEN) or non-targeting control siRNA (siCON) by immunoblotting 48 hours post-treatment. **(b)** Detection of FEN1 protein in XPF-proficient or XPF-deficient human cells treated with FEN1 siRNA (siFEN) or non-targeting control siRNA (siCON), as assessed by immunoblotting 48 hours post-treatment. The amount of FEN1 expression was quantified using ImageJ. The percentages shown below the gel were compared to the expression levels in control knockdown samples. This experiment was performed by Dr. Junhua Zhao. [Zhao, *et al.* (submitted *Nature Communications* 2016)].

3.4.6. Distinct mechanisms for replication-dependent and replication-independent H-DNA processing.

The interaction of NER and FEN1 proteins with H-DNA was confirmed by ChIP assays in yeast. Using primers flanking the human H-DNA sequences on the YAC (Table 3.1), we detected a 12-fold enrichment of Rad27(FEN1) at the H-DNA-forming region over that of a control region (Fig. 3.9a). Enrichment of Rad1(XPF) and Rad2(XPG) was also detected at the H-DNA region, at a 12-fold and 14-fold enrichment above the control region, respectively (Fig. 3.9a). Notably, the absence of Rad14(XPA) reduced Rad1(XPF) binding to H-DNA by ~4-fold (from 32-fold to 9-fold above the control, Fig. 3.9b), consistent with a requirement for functional NER to recruit ERCC1-XPF to H-DNA. In contrast, the enrichment of Rad27(FEN1) remained at similar level in the presence or absence of Rad14(XPA) (Fig. 3.9b). Taken together, these results support a role for NER in the recognition and processing of H-DNA *in vivo,* a mechanism distinct from that of FEN1.

DNA replication plays a key role in many types of DNA structure-induced mutation (254). Thus, we asked whether the effect of ERCC1-XPF and FEN1 on H-DNA mutagenesis was dependent on DNA replication. In both replication-proficient (Fig. 3.9c) and replication-deficient systems (Fig. 3.9d), H-DNA stimulated mutations over the background control levels. In both cases, XPF deficiency resulted in a significant decrease in H-DNA-induced mutagenesis

compared to XPF-proficient cells; however, its role in the replication-deficient system was more substantial than in the replication-proficient assay (Fig. 3.9c and d; *t*-test, *P* value <0.05). These results revealed a critical role for XPF in a replication-independent mechanism(s) of H-DNA-induced mutagenesis in human cells, consistent with our NER "damage" repair model (Fig. 1.7) (31,125,256). We posit that in non-replicating DNA, the "repair" cleavage of H-DNA by ERCC1-XPF and XPG results in DNA breaks that are processed by an error-prone microhomology-mediated end-joining (MMEJ) mechanism. In contrast, the effect of FEN1 on H-DNA-induced mutagenesis was dependent on DNA replication; FEN1-depletion in human cells resulted in an increase in H-DNA-induced mutagenesis only in replicating templates in both XPF-proficient and XPF-deficient cells (Fig. 3.9c and d). These results suggest that the role of FEN1 in H-DNA structure-induced mutagenesis was largely dependent on active DNA replication.

Because H-DNA can stall replication forks *in vivo* (322), we hypothesize that in the absence of FEN1, the H-DNA structures may persist at replication forks leading to fork collapse, which can result in DSBs and an increased frequency of mutagenic events. Thus, our data support a model in which FEN1 cleaves H-DNA structures (or their intermediates) during replication to resolve the mutagenic structures to ensure continuous replication. The resulting repair intermediates may then be processed by error-free repair pathways that are active during S-phase (Fig. 3.10).

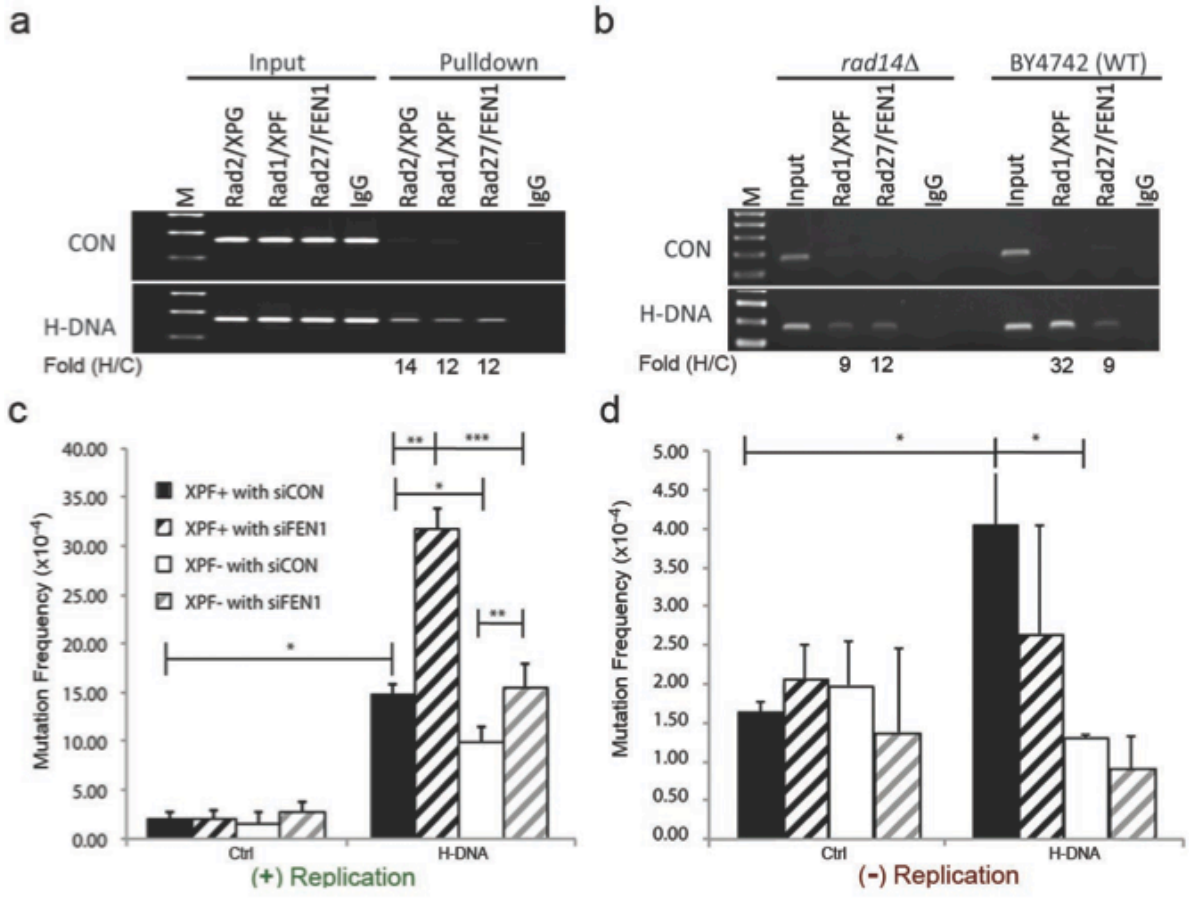


Figure 3.9. Association of endonucleases with H-DNA and effects of replication on H-DNA-induced mutagenesis. (a) Yeast ChIP results demonstrate enrichment of XPF(Rad1), XPG(Rad2), and FEN1(Rad27) at H-DNA relative to control B-DNA. The ChIP signal resulting from the PCR assay of the H-DNA-forming sequence was normalized to the control sequence for each antibody used for the pulldown; "Fold (H/C)". **(b)** ChIP results in *rad14Δ(XPA)* yeast demonstrate a requirement for Rad14(XPA) in the recruitment of XPF(Rad1), but not FEN1, to H-DNA. **(c)** and **(d)** Effects of XPF and FEN1 on H-DNA-induced mutagenesis in replicating (c) or non-replicating systems (d). Error bars represent s.d. from three repeats; two-sided *t*-test, *P* values <0.05 (*), <0.01 (**). These experiments were performed by Jennifer McKinney, Dr. Aklank Jain, and Dr. Junhua Zhao [Zhao, *et al.* (submitted *Nature Communications* 2016)].

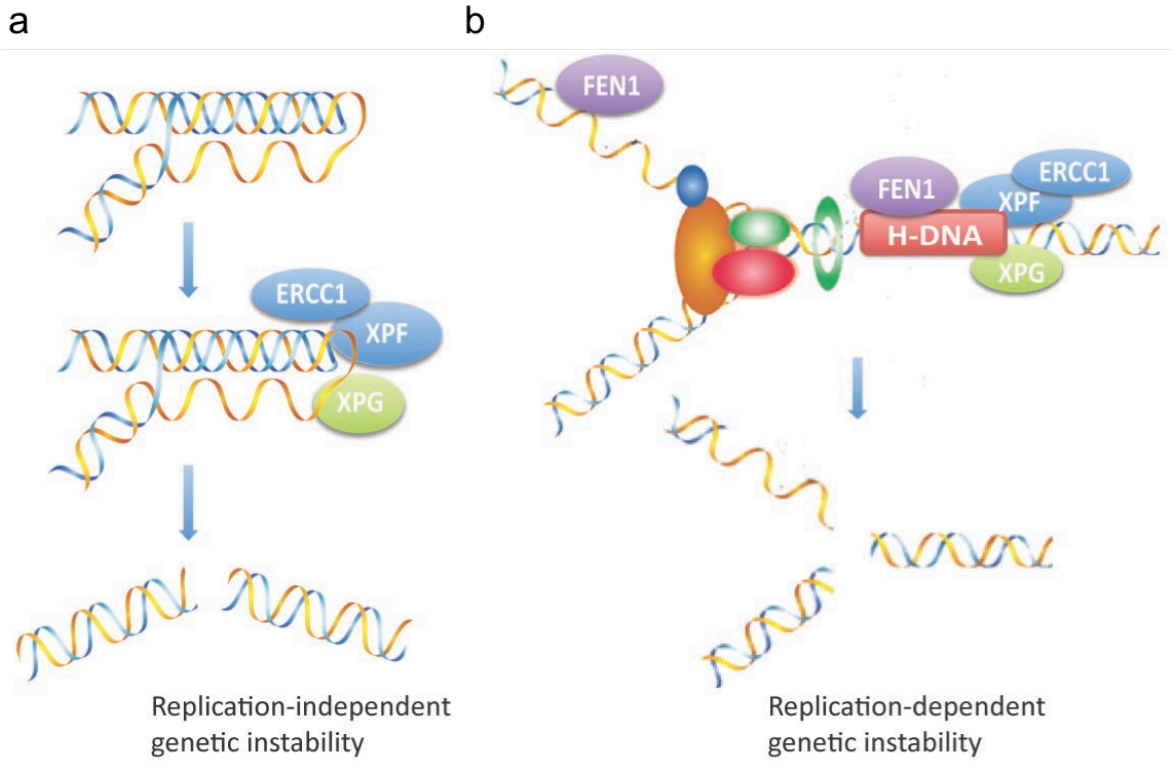


Figure 3.10. Models of replication-dependent and replication-independent H-DNA-induced genetic instability in eukaryotes. (a) Formation of the H-DNA structure is recognized as "damage" and is cleaved by NER nucleases, XPF/ERCC1 and XPG. This cleavage results in DNA strand breaks and subsequent deletions. **(b)** During replication, FEN1 cleavage of H-DNA can result in error-free processing. However, in the absence of FEN1, the H-DNA structure may persist at a replication fork, causing fork collapse, which could result in DSBs and subsequent deletions [Zhao, *et al.* (submitted *Nature Communications* 2016)].

3.5. Discussion

H-DNA structure-forming sequences often co-localize with chromosomal breakage and translocation hotspots in Burkitt's lymphoma, plasmacytoma, and leukemias (141,166,187,299,301). In various model systems, H-DNA structures have been shown to stimulate mutations/deletions and subsequent genetic instability (125,128). Our study identified two mechanisms of H-DNA-induced genomic instability; one replication-dependent, and the other replication-independent.

It has been reported that NER plays an important role in recognizing inter-molecular triplex structures (120,126,323). Cleavage of TFO-directed DNA interstrand crosslinks by NER proteins has also been detected (120,126,324-326). Due to the similarity in TFO-directed inter-molecular triplex DNA structures and intra-molecular triplex H-DNA structures, the potential role of DNA repair proteins on DNA structure-induced genetic instability has been speculated (256). Here, we identified the involvement of the Rad1(XPF) and Rad10(ERCC1) complex in H-DNA structure-induced genomic instability in yeast and detected cleavage of H-DNA by XPF in human cell extracts and purified XPG protein. In the NER mechanism the structure-specific nucleases, ERCC1-XPF and XPG, cleave at the 5' and 3' sides of bulky adducts, respectively, to excise the damaged strand (302,318,327,328). Our results demonstrating a role for ERCC1-XPF in H-DNA-induced DSB formation and the generation of deletions suggested

that H-DNA may be recognized as “damaged DNA”, where the H-DNA structures were cleaved by ERCC1-XPF and XPG, resulting in an error-prone abortive DNA repair via DSB intermediates (31,125,256). Due to the structural alterations found in H-DNA, the cleavage sites are possibly different from the canonical ERCC1-XPF and XPG substrates, which do not contain Hoogsteen H-bonding (303,318,329,330). XPG cleaved the triplex structure at the ssDNA and dsDNA junction of the 5’ overhang, consistent with previous reports (304,318,330). However, the predominant ERCC1-XPF cleavage site was identified in the loop 3’ to the dsDNA, and not at the dsDNA region (303). In any case, these results implicated H-DNA as a novel substrate for ERCC1-XPF and XPG cleavage, which may generate strand breaks for further repair processing.

DSBs near the H-DNA region were detected in human XPF-proficient cells but not in XPF-deficient cells, consistent with the cleavage of H-DNA by ERCC1-XPF and XPG. In fact, the majority of H-DNA-induced mutations generated in the human XPF- and XPG-proficient cells were deletions. In contrast, point mutations comprised the majority of deletions in the human XPF-deficient and XPG-deficient cells, indicating that ERCC1-XPF and XPG promote strand breaks at H-DNA structures. H-DNA structure-induced genomic instability was significantly reduced in XPF-deficient cells, and this effect did not require active DNA replication, suggesting that ERCC1-XPF may be functioning via NER in non-replicating cells. Further, the recruitment of XPF and XPG proteins to H-DNA required XPA in ChIP assays, further supported a role for NER in the mutagenic

processing of H-DNA. However, we cannot exclude a role for ERCC1-XPF outside of NER (303,331,332) or a role for XPG in transcription (333) as contributors to the variations in mutation frequencies and spectra identified.

In contrast to the effects of ERCC1-XPF on H-DNA-induced mutagenesis, which were independent of DNA replication, the stabilizing effect of FEN1 on H-DNA was largely dependent on active DNA replication. FEN1 plays an important role in 5' DNA flap removal during replication and replication fork rescue (305,306). The cleavage of the triplex structure with a 5' overhang by FEN1 was detected in our study and the depletion of FEN1 in human cells stimulated H-DNA-induced instability in a DNA replication-dependent manner. Thus, our results provide a plausible explanation for the processing of H-DNA during replication. In the absence of FEN1, the H-DNA structure may persist at a replication fork, causing fork collapse, which could result in DSBs and increased frequencies of mutagenic events, such as the deletions that we have observed.

Many more genes are likely involved in H-DNA-induced DNA breakage, repair and mutagenesis as a network in human cells, similar to that described in prokaryotes (334). Further work to identify candidate genes is warranted, as it will provide critical information for understanding the mechanisms of DNA structure-induced genetic instability in human disease.

3.6. Work in progress related to the H-DNA-induced mutagenesis project

We screened mutants deficient in MMR proteins to evaluate the effect of these proteins on H-DNA-induced genetic instability. While the H-DNA structure may mimic a bulky adduct and thus become a DNA “damage” substrate of the NER pathway, the structure contains a single-stranded loop that may also be recognized as “damage” by MMR components. Additionally, our lab has found that the MMR, MSH2-MSH3 recognition complex, and the MMR protein MLH1 are involved in the damage signaling and error-free repair of triplex-forming oligonucleotide (TFO) psoralen-directed inter-strand crosslinks (ICLs) (252,253). Moreover, our lab has shown interactions between MMR and NER proteins, e.g. MSH2-MSH3 and XPC, and XPA/RPA on TFO ICLs (130), suggesting that proteins from multiple repair pathways may be acting outside canonical roles in DNA damage recognition and repair to participate in H-DNA-induced mutagenesis or inhibition. Thus, we hypothesize that proteins from multiple DNA recognition/processing pathways are also likely to be involved in H-DNA-induced mutagenesis and it is of interest to screen for gene products involved in this processing in addition to those we identified above.

3.6.1. The role of MMR proteins in H-DNA-induced genomic instability in yeast.

We used the yeast deletion library to screen MMR components as described in the methods. Interestingly, we found that the *msh2* mutant strain resulted in a decrease in H-DNA-induced mutagenesis relative to background level (Fig. 3.11a; 1.3-fold induction compared to 9.4-fold in WT). The deletion of *msh3* did not result in a significant decrease in H-DNA-induced genetic instability (Fig 3.11a 10.3-fold compared to 9.4-fold in WT), while the deletion of other MMR components, *msh6Δ* and *mlh1Δ* did result in a reduction in H-DNA-induced mutation frequencies (Fig. 3.11a 3.6-fold and 4.7-fold, respectively compared to 9.4-fold in WT). Our results of an effect on H-DNA mutagenesis in *msh2Δ, msh6Δ,* and *mlhΔ,* but not *msh3Δ* suggest the Msh2-Msh6 complex recognizes H-DNA, which is interesting since its role in MMR is to recognize single base pair mismatches, while the Msh2-Msh3 complex recognizes up to 10 base pair loops (242,244,335-337); however, it should be noted that the elevated background genomic instability needs to be taken into consideration (Fig. 3.11a, compare blue bars). If the raw data is considered, it is comparable to that of WT (Fig. 3.11, compare red bars), demonstrating that our H-DNA-forming sequence is still mutagenic in these mutant strains.

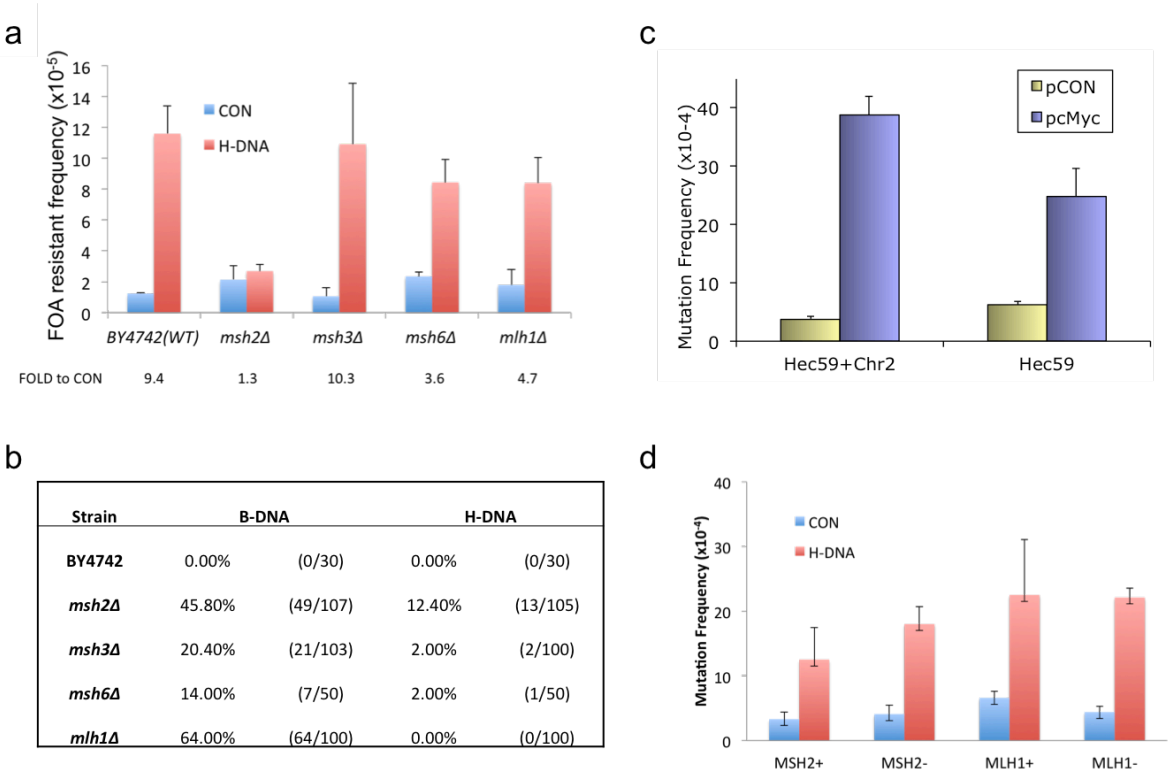
The analysis of the mutation spectra caused by H-DNA in *msh2Δ* revealed a large percentage of point mutations (Fig. 3.11b). Of the 105 FOA^R colonies we

analyzed in the H-DNA YAC in the *Δmsh2* strain, 13 maintained the right arm of the YAC, compared to 0 out of 30 in the WT strain. These results suggest that in the *msh2Δ* strain containing the H-DNA YAC, the FOA^R mutants are caused by random point mutations in the *URA3* gene, most likely due to the lack of functional MMR pathway, and as a result of H-DNA formation. A slight increase in point mutations in FOA^R colonies was also found in *msh3Δ* and *msh6Δ* strains containing the H-DNA YAC. These data are preliminary and statistical analyses still need to be assessed.

3.6.2. The role of MMR proteins in H-DNA-induced genomic instability in human cells.

We used the plasmid-based assay described in the methods and human MSH2-proficient and MSH2-deficient cell lines to evaluate the role of MSH2 in H-DNA-induced mutagenesis in human cells. Preliminary results in our lab demonstrated a decrease in H-DNA-induced mutagenesis in human MHS2-deficient cells compared to MSH2-proficient when compared to the control plasmid (Fig 3.11c, 4-fold in MSH2-deficient cells compared to 10.3-fold in MSH2-proficient cells), suggesting a role for Msh2 in promoting genetic instability caused by H-DNA, and validating our yeast data. However, my recent results suggested that there was no difference in H-DNA-induced mutagenesis in human

MHS2-proficient and MSH2-deficient cells (3.7-fold vs. 4.0-fold, respectively) as well as in MLH1-proficient and MLH1-deficient cells (6-fold vs. 4.0-fold, respectively) (Fig. 3.11d), suggesting that neither protein is involved in H-DNA-induced mutagenesis. These studies are on going and require further investigation to determine the role of MMR in genetic instability due to H-DNA structure formation.



Strain	B-DNA		H-DNA	
BY4742	0.00%	(0/30)	0.00%	(0/30)
msh2Δ	45.80%	(49/107)	12.40%	(13/105)
msh3Δ	20.40%	(21/103)	2.00%	(2/100)
msh6Δ	14.00%	(7/50)	2.00%	(1/50)
mlh1Δ	64.00%	(64/100)	0.00%	(0/100)

Figure 3.11. Preliminary data on the role of MMR proteins in H-DNA-induced mutagenesis in yeast and human cells. (a) H-DNA-induced FOA resistance on YACs in WT (BY4247) or MMR mutant strains. **(b)** Number of point mutations evaluated in FOA^R colonies for WT and MMR mutant yeast strains in control (B-DNA) and H-DNA YACs. **(c)** Previous data measuring H-DNA-induced mutation frequencies on *supF*-mutation reporters in WT (Hec59+Chr2) and MSH2-deficicient (Hec59) human cells. **(d)** Recent data measuring H-DNA-induced mutation frequencies on *supF*-mutation reporters in MSH2 and MLH1-proficient and deficient human cell lines. Experiments in a and b were performed by Dr. Junhua Zhao, c was performed by Ms. Fei Yang and d were performed by Jennifer McKinney.

Chapter 4: Stable Integration of ADAR1 Over-expression Vector in COS-7 Mammalian Cell Line

4.1. Abstract

More than 10 different non-B DNA structures have been described that differ from the well-known Watson-Crick B-DNA double helix. For example, Z-DNA is a left-handed helix that forms at sequences of alternative purines and pyrimidines with the energy provided by negative super-coiling, resulting from an open B-DNA duplex. Sequences that can form Z-DNA are found in various species of prokaryotes and eukaryotes, and are very abundant in the human genome, occurring 1/3,000 base pairs. While Z-DNA has been found to play a role in gene regulation, we have demonstrated that Z-DNA is mutagenic, resulting in double-strand breaks (DSBs) in human cells. Evidence suggests that Z-DNA is involved in several important biological processes, and Z-DNA-binding proteins that specifically bind to the Z-DNA structure have been identified. Adenosine deaminase acting on RNA 1 (ADAR1-p150) is a human protein that contains Z-DNA-binding domains and further suggests that Z-DNA is involved in transcription. We have obtained a plasmid vector that over-expresses ADAR-p150; a tool that we will use to further study the role of Z-DNA-induced mutagenesis. Here we describe the development of a COS-7 mammalian cell line in which the ADAR1-p150 over-expression vector has been stably integrated.

4.2. Introduction

Genomic DNA typically exists in the canonical, right-handed B-conformation that was described in 1953 by James Watson and Francis Crick (Fig. 4.1a, “B-DNA”) (1). However, DNA is a dynamic molecule, and can adopt several secondary conformations, collectively referred to as non-B DNA structures, that differ significantly from that of B-DNA. More than 12 non-B DNA structures have been identified and described, including the left-handed Z-DNA conformation, which is the focus of this chapter (Fig. 4.1b) (10-12,24,77,338). Sequences that can adopt non-B DNA structures are naturally occurring and abundant in the human genome. For example, sequences capable of adopting a Z-DNA structure have been found to occur ~1/3000 bp (41). The structure of Z-DNA was first described in 1979 and made the cover of *Nature* (8), and continues to be an active area of research.

Z-DNA can form at sequences of alternating purines and pyrimidines (i.e. $(CG)_n$ or $(AT)_n$) repeats, resulting in a left-handed double-helical structure. Contrary to B-DNA, the purines within the Z-DNA region are reversed and positioned over the sugar in a *syn* conformation, while the pyrimidines remain in the *anti* conformation [reviewed in (22)]. In order to maintain normal Watson Crick base-pairing, the sugar pucker is altered from the C2- to the C3-endo position, resulting in the characteristic zigzag pattern of the sugar-phosphate backbone, for which is was named (339).

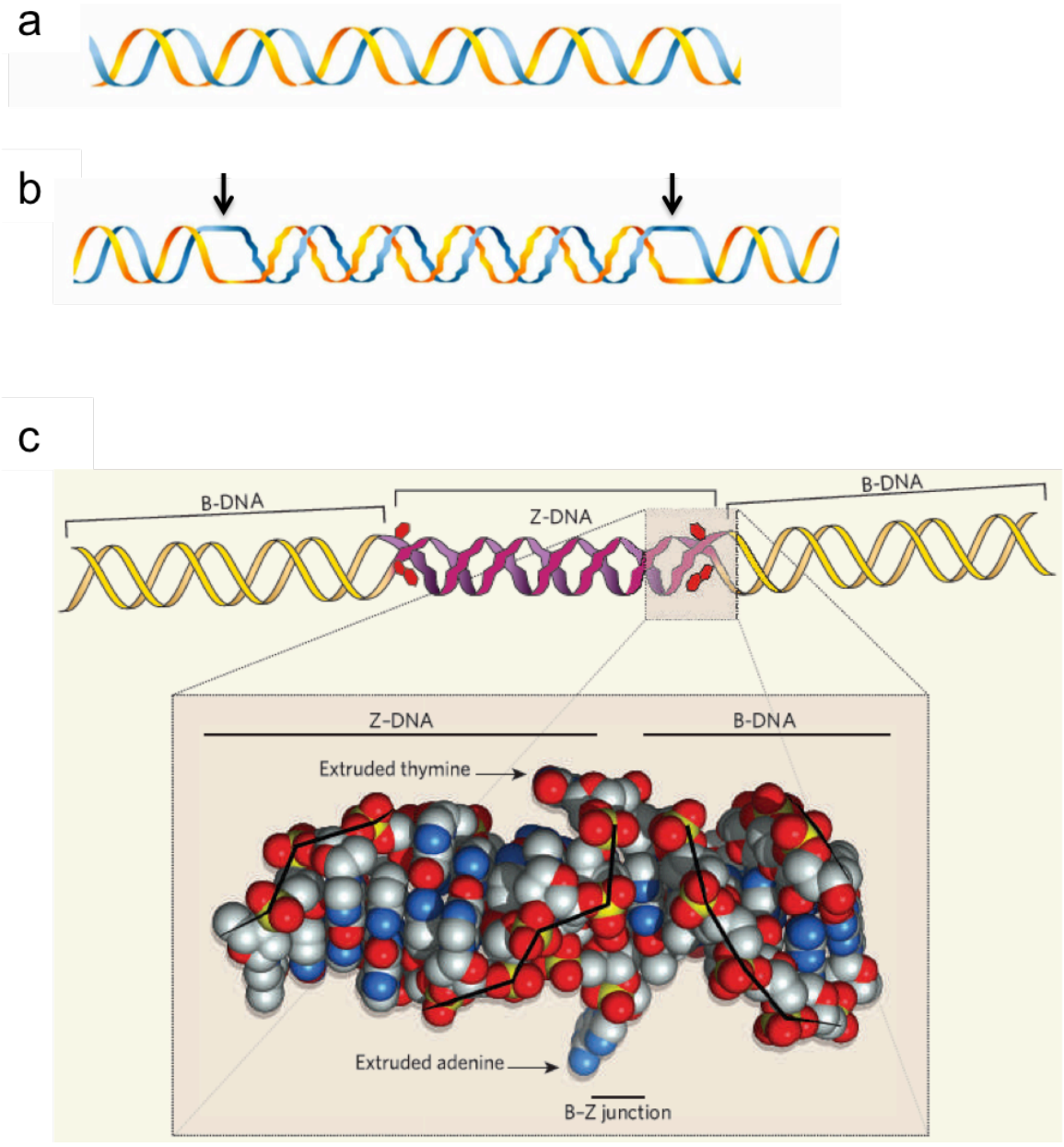


Figure 4.1. DNA structures. (a) Watson-Crick double-stranded B-DNA. **(b)** Left-handed Z-DNA structure shown the center of double-stranded B-DNA. Arrows indicate single-stranded B-Z junctions [a and b adapted from (256)]. **(c)** Van der Waals view of B-Z junction showing extruded base pair as determined by crystal structure [adapted from (340)].

Significant structural alterations to the DNA double helix occur in Z-DNA regions, including loss of the major groove and a more compact helix with 12 bp/turn compared to the 10.5 bp/turn of B-DNA (8,25). Additionally, the crystal structure of B-Z junctions revealed an extruded base pair, resulting in a small single-stranded region at the junction (Fig. 4.1c) (18).

Non-B DNA-forming sequences occur at high rates in the genome and are found in both prokaryotes as well as evolutionarily conserved across multiple species of eukaryotes (24). Thus, the biological function of such structures has been widely studied and evidence has shown that non-B DNA structures are likely involved in DNA metabolic processes, such as replication and transcription (11,32,60,123). Z-DNA-forming sequences map to transcription start sites, and it has been shown that this structure can play a role in regulation of gene expression (49,50,255,257). Non-B DNA-forming sequences can also contribute to genomic instability, which has been studied extensively in our laboratory; however, specific mechanisms have yet to be fully elucidated (see Chapters 1, 2, and 3 for more on the role of non-B DNA in mutagenesis and disease).

Several non-B structures exhibit a number of similar characteristics (i.e. structural aspects, location, requirement for negative supercoiling, etc.); however, one feature unique to Z-DNA is the existence of several proteins that have been identified to contain Z-DNA-binding domains (ZBDs). These motifs have been shown to specifically recognize a left-handed, double-stranded helical structure via a winged-helix conformation. Z-DNA-binding proteins (ZBPs) have been

identified in viral (62-64), bacterial (66,67), yeast (68) and human cells (67,74,75,341), among others (69-74,342). ADAR1, found in humans, is among those ZBPs that is the most extensively studied, and is the focus of this chapter.

Adenosine deaminase acting on RNA 1 (ADAR1), a member of the deaminase family, modifies viral and cellular double-stranded RNA (dsRNA), converting adenosine to inosine during A-to-I RNA editing (343-347) (Fig. 4.2a). During translation and RNA replication inosine is recognized as guanine and base pairs with cytosine instead of uracil in dsRNA, resulting in base substitution that can alter the sequence information and structure (348,349). Other proteins exist in the ADAR family (ADAR2 and ADAR3); however, ADAR1 is the only member to have ZBDs (Z-α and Z-β) (350).

Separate promoters and alternative splicing of the *ADAR1* gene give rise to two isoforms of ADAR1 (350-353) (Fig. 4.2b). The constitutively expressed ADAR1-p110 has a truncated N-terminus, and is found exclusively in the nucleus. Although ADAR1-p110 contains a single ZBD (Z-β), it only binds to dsRNA, as the Z-β domain alone is not sufficient to bind to Z-DNA. The full-length protein, ADAR1-p150, is found in the mainly in the cytoplasm, is interferon (IFN) inducible, and contains both Z-α and Z-β ZBD; thus it binds dsRNA and Z-DNA (352,354,355). The Z-α domain has been shown to specifically recognize the zigzag conformation and the *syn* conformation of the purine residues in Z-DNA, when compared to similar sequences that do not adopt the Z-DNA structure (351,356).

The biological role for Z-DNA interaction with ADAR1 has not been clearly defined; however, one postulation is that Z-DNA formation is induced during transcription behind an active RNA polymerase, which subsequently binds ADAR1, and serves to localize the mRNA editing activity of ADAR1 to a specific region of the gene (25) (Fig. 4.2c). This interaction between Z-DNA and ADAR1 may also allow editing to occur as the nascent mRNA is made before other modifications occur, as well as transiently preventing further transcription until mRNA editing is completed (25).

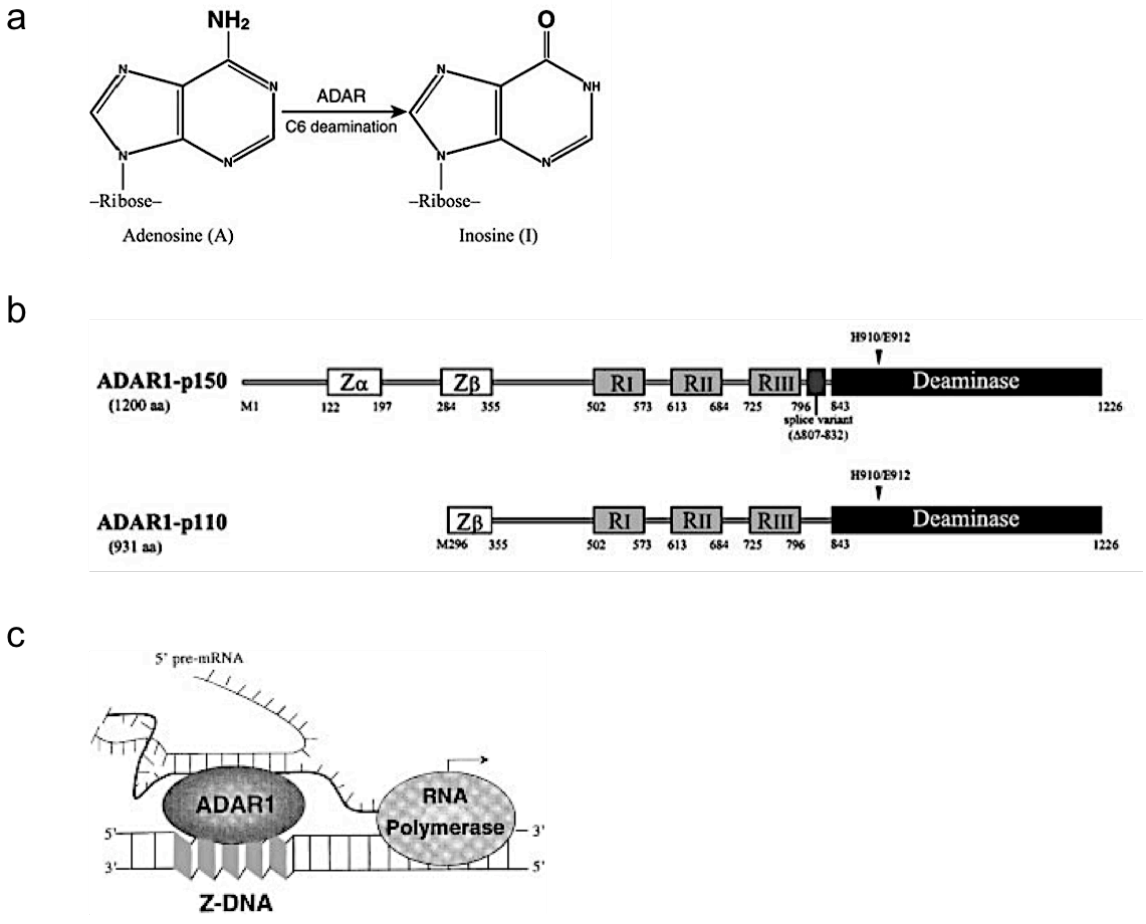


Figure 4.2. Adenosine deaminase acting on RNA 1 (ADAR1). (a) ADAR1 converts adenosine to inosine via a C6 hydrolytic deamination reaction [adapted from (357)]. Inosine will be recognized as guanine during translation, resulting in mutation. **(b)** ADAR1-p150 and p110 isoforms depicting organization of Z-DNA-binding domains (Z-α and Z-β), dsRNA-binding domains (RI-RIII), and the deaminase catalytic domain. Arrowheads indicate amino acid residues that are important for catalytic function and numbers directly below each schematic represent amino acid position of each domain [adapted from (357)]. **(c)** Model for regulation of ADAR1 activity by Z-DNA as proposed by Alan Herbert and Alexander Rich in 1999 (25). Negative supercoiling that occurs behind a polymerase promotes formation of Z-DNA. ADAR1 binds Z-DNA, which activates editing of the nascent mRNA as the transcript forms [adapted from (25)].

The existence of Z-DNA-binding proteins suggest a biological function for Z-DNA in several metabolic process. Importantly, sequences capable of forming Z-DNA are enriched at genomic “hotspots” that are prone to chromosomal breakage and translocations related to human disease (122,123,147,163,165,166,258). We, and others, have found that Z-DNA is mutagenic in bacteria, yeast, and mammalian cells, and results in the formation of DSBs (127,128,168); however the mechanism behind Z-DNA-induced mutagenesis has yet to be fully elucidated. Z-DNA-binding proteins, such as ADAR1 provide a useful tool to expand our knowledge of biological roles of Z-DNA, both advantageous and deleterious.

Here we stably integrated an expression vector that over-expresses a FLAG-tagged ADAR1-p150 isoform into mammalian COS-7 cells and detected a 5.8-fold average increase in expression levels of ADAR1-p150 over endogenous levels. This stable cell line will be used for further studies to determine the role of the Z-DNA-binding protein, ADAR1-p150, in Z-DNA-induced genomic instability in mammalian cells.

4.3. Materials and Methods

4.3.1. Cell lines and plasmids. The pCMV-FLAG-ADAR1-p150 expression vector containing a FLAG-tagged ADAR1-p150 gene inserted in the multiple cloning site (MSC) between *Hind*III and *Xba*I in the parent plasmid pcDNA3 (Invitrogen™, Life Technologies), resulting in over-expression of the p150 isoform of FLAG-tagged ADAR1 was a gift from Dr. Qingde Wang at The University of Pittsburgh, PA (358). An empty vector with the same parental plasmid backbone (pcDNA3.1 (+/-)) as the ADAR1 over-expression vector (Invitrogen™, Life Technologies) was used as a control. Mammalian COS-7 (ATCC) cells were used to make stable cell lines over-expressing ADAR1 and were maintained in maintained in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, CA) with 10% fetal bovine serum (FBS) and 400-800 µg/mL G418•Sulfate (Enzo Life Sciences, Farmingdale, NY). *E. coli* MBM7070 cells (F-*lacZ (am)CA7020, lacY1, hsdR−, hsdM+, araD139 Δ(araABC-leu)7679, galU, galK, rpsL, thi*) (Lucigen, Middleton, WI) (316) were used for plasmid transformation.

4.3.2. Verification of ADAR-p150 expression vector. The pCMV-FLAG-ADAR1-p150 expression vector or pcDNA3.1 control vector were transformed into MBM7070 *E. coli* cells and a Qiagen HiSpeed Plasmid Maxi Kit was used to prepare large-scale amounts of the plasmid vectors. The pCMV-FLAG-ADAR1-

p150 expression vector (250 ng) was digested with *Hind*III and *Xba*I to verify that the FLAG-tagged ADAR1 insert was present, and sent for direct DNA sequencing with primer (5'-CGT GTA CGG TGG GAG GTC TA-3') to confirm the presence of the FLAG-tag.

4.3.3. Development of stably integrated cell lines. A large-scale digest (8 µg plasmid DNA) was performed on the pCMV-FLAG-ADAR1-p150 and pcDNA3.1 control vectors using *Afl*III to linearized the plasmids and release a 2-kb fragment containing the Ampicillin-resistance gene. Reactions were performed at 37°C overnight in 100 µL total volume, and linearized products were separated on a 1% agarose gel. The 6-kb fragment containing the ADAR1-p150 gene and the 3.5-kb fragment from the control plasmid were excised from the gel and purified using a Qiagen QIAquick Gel Extraction Kit. Ethanol precipitation was performed on isolated DNA in 100% ethanol at -20°C overnight. DNA was pelleted by centrifugation, washed with 70% ethanol, again pelleted by centrifugation, and resuspended in 10 µL sterile water. The linearized and purified DNA fragments were transfected into COS-7 cells by electroporation using a Nucleofector (Lonza, Walkersville, MD). Transfected cells were maintained in DMEM + 10% FBS without antibiotic selection for 5 days, followed by 800 µg/mL G418•Sulfate treatment for Neomycin selection of cells containing incorporated plasmid DNA until single colonies were visible, then cells were maintained in DMEM + 10%

FBS with 400 µg/mL. Cells pellets were harvested for confirmation of ADAR1-p150 over-expression vector by Western blotting.

4.3.4. Western blotting. Cells were lysed in ice-cold RIPA buffer (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10 mM DTT, 0.1% Triton X-100 and Complete™ proteinase inhibitor cocktail (PIC) (Roche)) for 30 min at 4°C with rotation, followed by sonication at 40 Amps for 5 repeats of 30 second sonication, 30 second incubation on ice using an EpiSonic™ multi-functional bioprocessor (Epigentek, Farmingdale, NY). Samples were centrifuged at 4°C for 20 min at 12,000 rpm, and the supernatant was collected for further analysis. A BCA assay was performed to determine protein concentrations of the lysates. Cell lysates (20-30 µg) were mixed with SDS gel-loading buffer and heated at 95°C for 10 min, separated by gel electrophoresis on a 10% SDS-polyacrylamide gel, and transferred to nitrocellulose membranes (Bio-Rad Laboratories, Inc., Hercules, CA). Purified amino-terminal FLAG-BAP™ Fusion Protein was used as a control (20-80 ng) (Cat. Num. P7582; Sigma-Aldrich, St. Louis, MO). The blots were blocked for 1 hour at room temperature in TBST buffer [tris buffer saline (TBS)] containing 5% nonfat milk and 0.1% Tween 20). The blots were then incubated with diluted primary antibody for 1-2 hours at room temperature. Primary antibodies used in this study included rabbit anti-human PCNA antibody (Santa Cruz Biotechnology, Santa Cruz, CA), monoclonal ANTI-FLAG® M2, Clone, M2 (Cat. Num. F1804; Sigma-Aldrich, St. Louis, MO), and ADAR purified MaxPab

mouse polyclonal anti-human antibody (Cat. Num. B01P; Abnova, Taipei, Taiwan). The blots were washed three times with TBST buffer for 5 min at room temperature, and once with TBS buffer for 5 min at room temperature. The blots were incubated for 1 hour at room temperature with horseradish peroxidase-conjugated anti-rabbit IgG or mouse IgG secondary antibodies (Bio-Rad Laboratories, Hercules, CA). After three washes with TBST buffer for 10 min, and one wash with TBS buffer for 5 min at room temperature, bound secondary antibody was detected using Amersham ECL-Plus Western Blotting Detection Reagent (GE Healthcare Life Sciences). Blots were exposed to Kodak Biomax XAR film, and developed. ADAR1 expression was quantified and normalized to respective PCNA expression and ADAR1 expression in untreated cells using NIH ImageJ64 software. Western blotting was repeated four times. Significance was determined as a p value of <0.05 using Student's t-test.

4.4. Results

4.4.1. Development of a mammalian cell line with stably integrated pCMV-ADAR1-p150 over-expression vector.

Following the preparation of a large-scale quantity of the plasmid received from Dr. Qingde Wang, we verified the presence of the ADAR1-p150 insert by restriction digest with *Hind*III and *Xba*I to release the ADAR1-p150 fragment (Fig. 4.3a, red lines). Products were separated on an agarose gel to verify fragment size (Fig. 4.3b). Verified pCMV-ADAR1-p150 and pcDNA3.1 control vectors were digested with *Afl*III to linearize and release the ampicillin-resistance gene (~2 kb-fragment) (Fig. 4.3a, blue lines), and fragments were separated by gel electrophoresis (Fig. 4.3c). Remaining fragments of linearized DNA containing the FLAG-ADAR1-p150 gene (~7 kb) and control DNA (~3 kb) were isolated. Purified DNA fragments were transfected into mammalian COS-7 cells and selected for integration by neomycin-resistance with G418•Sulfate for several weeks.

Stable integration of the FLAG-ADAR1-p150 expression vector was verified by Western blotting. Our results revealed that untreated cells contained the same level of the ADAR1-p150 protein expression as the cells transfected with the control vector, while the cells transfected with the FLAG-ADAR1-p150 expression vector resulted in over-expression of ADAR1-p150 protein (average

of 5.8-fold above untreated cells vs. 3.9-fold above pcDNA3.1 transfected cells) (Fig. 4.3d). We expected some variability between ADAR1-p150 measurements because it is not constitutively expressed; however we demonstrated a significant increase in ADAR1-p150 expression in the cells transfected with the FLAG-ADAR-p150 vector ($p<0.05$ when comparing 5.8-fold vs. 3.9-fold, Student's t-test) (Fig. 4.3d, lower panel). Because the ADAR antibody is polyclonal and able to detect the ADAR1-p110 isoform expression as well, we used this isoform expression level as a control to definitively confirm that only the Z-DNA-specific ADAR1-p150 isoform was over-expressed within our system. The ADAR1-p110 isoform expression remained the same (1.24-fold above untreated cells vs. 1.12-fold above pcDNA3.1 transfected cells (Fig. 4.3d, lower panel), confirming that we had developed a stably integrated system that specifically over-expressed the Z-DNA-binding protein, ADAR1-p150. Within the FLAG-ADAR1-p150 over-expression cell line, there was a significant increase in the p150 isoform over p110, whereas no such increase was seen in the pcDNA3.1-transfected cells (4.8-fold vs. 1.2-fold, respectively) (Fig. 4.3d, lower panel).

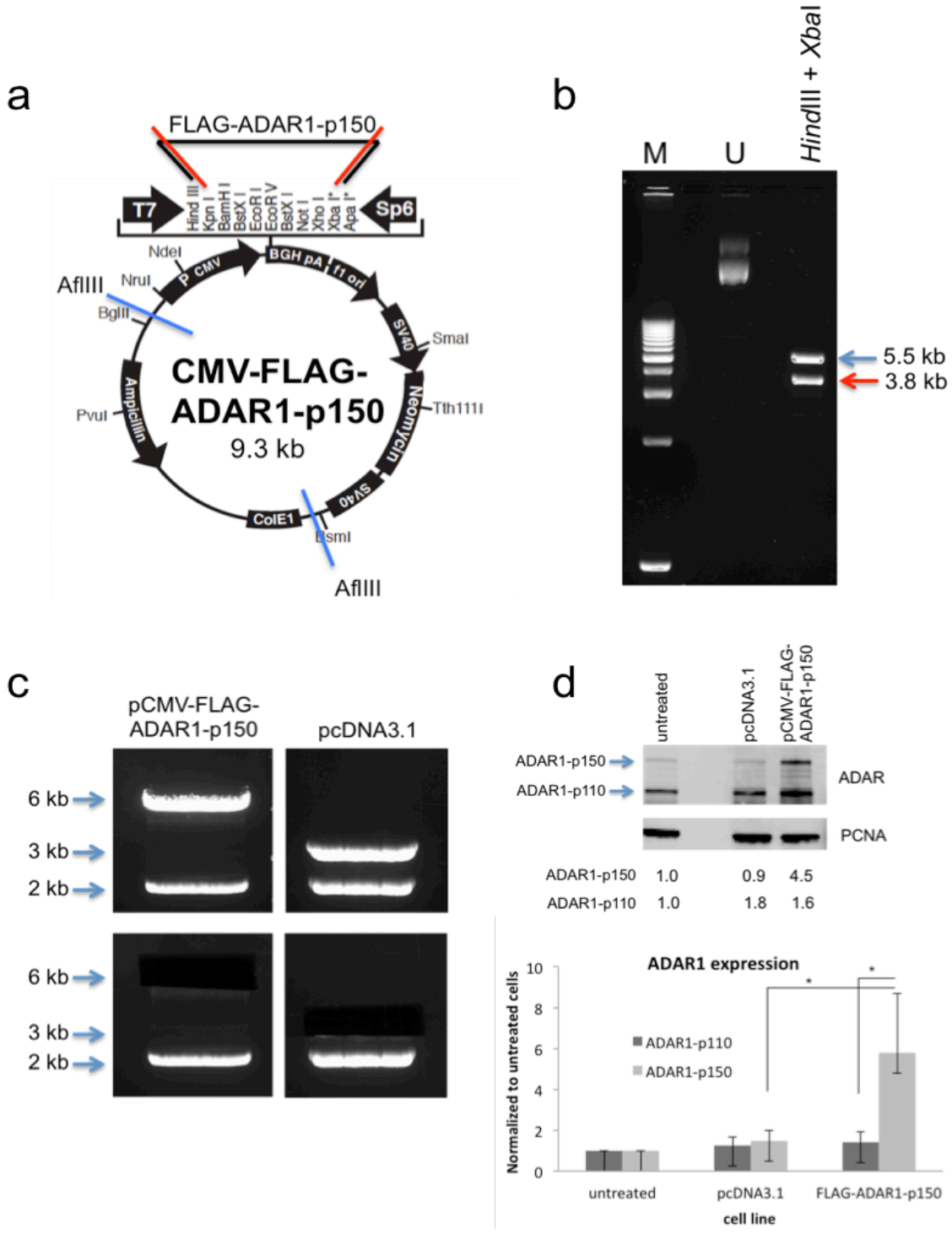


Figure 4.3. Verification and establishment of stable mammalian cell line over-expressing FLAG-ADAR1-p150.

Figure 4.3. Verification and establishment of stable mammalian cell line over-expressing FLAG-ADAR1-p150. (a) pCMV-FLAG-ADAR1-p150 expression vector with the FLAG-ADAR1-p150 gene inserted between the *Hind*III and *Xba*I restriction sites (red lines). Blue lines indicate the location of the *Afl*III digest to linearize the plasmid, releasing a 2-kb fragment containing the Amp^R gene for integration into mammalian COS-7 cells (modified from an Invitrogen plasmid map). **(b)** Confirmation of the correct FLAG-ADAR1-p150 insert following large-scale DNA preparation. Isolated DNA was digested with *Hind*III and *Xba*I restriction enzymes, resulting in two fragments; the insert (3.8 kb), and the backbone (5.5 kb). Undigested plasmid DNA (U) and a 1-kb marker (M) were included as controls. **(c)** pCMV-FLAG-ADAR1-p150 and pcDNA3.1 plasmids were linearized by restriction digestion with *Afl*III to release a 2-kb fragment containing the Amp^R gene. Fragments were separated by gel electrophoresis (top panels), and the larger fragment was excised and isolated from the gel (bottom panels). **(d)** Stable integration of linearized pCMV-FLAG-ADAR1-p150 expression vector into the COS-7 genome was verified by Western blotting using polyclonal ADAR antibody to detect ADAR1 expression. PCNA was used as a loading control, and untreated cells and cells transfected with the pcDNA3.1 control vector were included as negative controls. Protein expression was normalized to PCNA expression, then to untreated cells. Shown is a representative blot of multiple repeats and average of four repeats below (* denotes p value <0.05). These experiments were performed by Jennifer McKinney.

4.5. Discussion and Future Directions

We have developed a mammalian cell line in COS-7 cells that specifically over-expressed the ADAR1-p150 protein ~4.5-fold over untreated cells, and ~4.9-fold over cells transfected with the control vector. While there are no commercially available antibodies that are specific to the ADAR1 p150 isoform, we have found a reliable antibody that can detect over-expression of the ADAR1-p150 compared to untreated cells. Initially, having the FLAG-tag would have allowed the use of a FLAG antibody rather than a polyclonal ADAR antibody to detect over-expression; however, several attempts to use a FLAG antibody were unsuccessful. Thus, we determined that the ADAR polyclonal antibody was able to detect sufficient over-expression when compared to the expression level of the alternate non-Z-DNA-binding protein, ADAR1-p110.

4.5.1. Immediate goals using the established cell lines described in this chapter.

Coinciding with the main research interest of our laboratory focusing on the relationship between non-B DNA and genomic instability, we will use the cell lines established in this study to determine the mutagenic effect of Z-DNA-forming and control plasmids in untreated, pcDNA3.1 control, and FLAG-ADAR1-p150 over-expressing cells lines. We will transfect Z-DNA-forming and control

mutation-reporter plasmids into these cell lines and perform mutagenesis assays including mutation frequency and spectra assays as described in Chapters 2 and 3 and previously (262). Whether or not we detect a difference in Z-DNA-induced mutagenesis with and without over-expression of ADAR1-p150 will be interesting and provide information as to the potential role of Z-DNA-binding proteins in genomic stability. Our prediction is that ADAR1-p150 will bind Z-DNA and stabilize this mutagenic structure, and will therefore lead to an increase in Z-DNA-induced mutagenesis.

4.5.2. Long-term goals using the established cell lines described in this chapter.

If results from the mutagenesis assays suggest a role for ADAR1-p150 in Z-DNA-induced genomic instability, we would also perform Chromatin Immunoprecipitation (ChIP) assays to further confirm association of ADAR1-p150 at the Z-DNA-forming region. We will use purified ADAR1-p150 protein with control and Z-DNA-forming plasmids to perform electrophoretic mobility shift assays (EMSA) and confirm binding as well. We will further explore ADAR1-p150 in DNA repair deficient cells (such as those described in Chapters 2 and 3), as well as how this protein effects Z-DNA-induced mutagenesis in the presence of exogenous DNA damaging agents, such as UV irradiation. We will also transiently deplete ADAR1-p150 by siRNA knock-down in COS-7 cells and

perform mutagenesis assays as described above with control and Z-DNA-forming mutation-reporter plasmids to further characterize the role of ADAR1-p150 in Z-DNA-induced mutagenesis. If ADAR1 does not effect Z-DNA-induced mutagenesis, then we can explore other known ZBPs, such as the tumor-associated protein, DLM-1 [also known as Z-DNA-binding protein 1 (ZBP1)] or DNA-dependent regulator of IFN-regulatory factors (DAI), for an effect on Z-DNA-induced mutagenesis.

Chapter 5: Future Directions, Summary, and Significance

5.1. Future Directions

Here, we have demonstrated that repair proteins from NER and MMR pathways are involved in Z-DNA and H-DNA-induced mutagenesis in yeast and human cells. Thus far, proteins from both pathways were found to be involved in Z-DNA-induced mutagenesis, while H-DNA-related mutations required several proteins from the NER pathway, suggesting that functional NER was required rather than an interaction among proteins from the two repair mechanisms. The mechanism(s) responsible for genomic instability caused by non-B DNA structures is clearly complex and many more genes are likely to be involved in breakage, repair and mutagenesis as a network in human cells, similar to that described in prokaryotes (334). Further studies to identify candidate genes and more concisely determine mechanistic steps are warranted, as it will provide critical information for understanding the mechanisms of DNA structure-induced genetic instability in human disease. These studies are outlined below.

5.2. Immediate Goals

5.2.1. Determine the extent to which the Z-DNA-binding protein, ADAR1 has an effect on Z-DNA-mutagenesis.

We have developed a stably transfected mammalian cell line that over-expresses the Z-DNA-binding protein (ZBP), ADAR1 (see Chapter 4). We will perform mutagenesis and mutation spectra analyses similar to those discussed in Chapter 2 using Z-DNA-forming and B-DNA-forming control plasmids in these cells to determine if the induction we see in WT cells is effected by ADAR1. We predict that ADAR1 will bind to and stabilize Z-DNA, resulting in an increased mutation frequency compared to cells containing endogenous levels of ADAR1; however, the binding of ADAR1 could possibly prevent binding and processing of repair proteins, which may decrease mutations near the Z-DNA site. Conversely, we will deplete ADAR1 to further determine the effect of ADAR1 on Z-DNA-induced mutagenesis. As discussed in detail in Chapter 4, we will also perform ChIP and/or gel mobility-shift assays to demonstrate binding of ADAR1 to our Z-DNA-forming plasmid. Additionally, we could test other ZBPs, such as DAI. The results we obtain from these assays will provide important knowledge in the role of non-B DNA-binding proteins/ligands in preventing or promoting genetic instability.

5.2.2. Further characterize the relationship between repair proteins found to be involved in Z-DNA-induced mutagenesis.

We have demonstrated that the MMR and NER repair proteins MSH2-MSH3 and (Rad10-Rad1)ERCC1-XPF, respectively, are involved in Z-DNA-induced mutagenesis in yeast and human cells. We will further characterize this relationship that appears to function outside of their canonical roles in the NER and MMR pathways, as other components of these pathways did not demonstrate an effect on Z-DNA-induced mutagenesis in our assays. As discussed in Chapter 2, MSH2-MSH3 and (Rad10-Rad1)ERCC1-XPF have previously been shown to interact in yeast and human cells (280-298), and it was found that a region of ERCC1 that partially overlapped with XPF-binding domain was required for interaction with MSH2 (296). We will determine if this region of ERCC1 is also involved in the interaction we found between MSH2-MSH3 and ERCC1-XPF by using similar truncated ERCC1 proteins and co-immunoprecipitation or ChIP assays in human cell lines, such as those used in the studies presented in Chapter 2. Additionally, we could test truncated MSH2, MSH3, or XPF as it not known which components of the complexes are involved in the interaction. We can perform these experiments in the presence or absence of Z-DNA-forming or control, non-B DNA-forming plasmids to confirm the interaction is specific to Z-DNA.

5.2.3. Double-knock out experiments in yeast and/or human cells.

Since multiple proteins from within the same or different pathways are involved in Z-DNA-induced mutagenesis, testing one gene at a time may not fully reflect the effect of a protein on mutagenic processing of Z-DNA. Thus, we could first construct yeast knockout strains [using previously described methods (359)] that are deficient in two or more genes and perform the experiments described in Chapter 2 to more fully determine the function of those genes in processing Z-DNA. Examples include *rad1Δ msh2Δ*, *rad10Δ msh2Δ*, and *rad1Δ msh3Δ* double knockouts. If warranted, we could also perform mutagenesis assays in human cells that are deficient in both XPF and MSH2 by transiently depleting MSH2 in XPF-deficient cells lines, to determine if there is an even greater effect on Z-DNA-induced mutagenesis than either single deficiency alone. We could follow up these experiments, by performing ChIP assays to determine if other proteins may be enriched at the Z-DNA-forming site in the absence of XPF and MSH2. Additionally, we could repeat biochemical experiments described in Chapter 2 to determine if cleavage near the Z-DNA-forming site by XPF is effected by the loss of MSH2.

5.2.4. Determine the extent to which longer repeats capable of forming multiple non-B DNA structures effect mutagenesis in yeast.

The yeast experiments described in this document use short mirror-repeat or $(CG)_{14}$ repeat sequences that are capable of forming H-DNA or Z-DNA, respectively; however, in humans, non-B DNA-forming sequences are often much longer, and may contain sequences capable of forming more than one type of non-B DNA structures. We have a longer non-B-DNA insert, as well as a longer (~500 bp) control insert on a YAC that we will also use to screen repair-deficient mutants as described in Chapters 2 and 3. We found that different mechanisms appear to be involved in mutagenesis caused by H-DNA or Z-DNA in yeast; thus, it would be particularly interesting to see which gene products are found to be important in mutagenesis stimulated by this longer insert that is capable of forming multiple structures. In this way, these studies would not be specific to H-DNA or Z-DNA because the long insert is capable of forming multiple non-B DNA structures, but they would serve as a general non-B DNA-induced mutagenesis assay in yeast cells. Ongoing studies in the lab are aimed at determining which structure might form in an overlapping sequence that is capable of forming multiple structures under various conditions (i.e., a sequence that can form H-DNA as well as G4 DNA). We are studying what, if any, effect the formation of one structure might have over the formation of the other, and

whether the likelihood of one structure to form over another has implications on genomic instability.

5.2.5. Identify additional genes involved in stimulation or inhibition of non-B DNA-induced mutagenesis.

In the studies described in Chapters 2 and 3, we identified important components of Z-DNA and H-DNA-induced mutagenesis. However, it is clear that this mechanism is complex and there other components are likely to be involved, e.g. other structure-specific nucleases. The yeast knockout library we have will allow us to screen many more gene products for involvement in non-B DNA-induced mutagenesis in an unbiased fashion. Our previous results suggested that H-DNA and Z-DNA-induced mutagenesis in humans may be due to non-homologous end-joining (NHEJ) or micro homology-mediated end-joining (MMEJ) mechanisms (125,127,171); therefore we will screen yeast strains deficient in major components of these mechanisms, including Dnl4 and Lif1 (yeast homologues to human DNA ligase IV and XRCC4, respectively), which are responsible for the ligation step in end-joining, to determine if these gene products also play a role in non-B DNA-induced mutagenesis.

Moreover, DNA helicases, such as BLM, WRN, DHX9, and FANCJ have been implicated in maintaining genomic integrity (360-363), and human diseases caused by deficiencies in these enzymes are characterized by genomic instability

and increased susceptibility to cancers (364,365). We, and others, have demonstrated the ability of these helicases to unwind non-B DNA structures (131,172,366-370). For example, we found that human DHX9 recognizes and binds to H-DNA, and protects against H-DNA-induced mutagenesis, suggesting the helicase unwinds H-DNA to protect genomic integrity (172). Therefore, it is likely that helicases may also unwind other non-B DNA structures, such as Z-DNA, and we will test this hypothesis by performing mutagenesis and *in vitro* binding assays described in Chapters 2 and 3, using a Z-DNA-forming and control plasmid in DHX9-defective human cells, similar to the studies performed with H-DNA (172). If warranted, we will test other helicases for possible activity to process Z-DNA. These studies would provide additional valuable information on how cells counteract the potential genetic instability brought on by the formation of non-B DNA structures, and if this counteraction is structure-specific.

Similarly, topoisomerases may have a destabilizing effect on Z-DNA. Topoisomerase I and II can alleviate the stress that negative supercoiling exerts on DNA by cleaving one or both strands, unwinding, and re-annealing the cut strands [reviewed in (371)]. The level of Z-DNA formation has been shown to correlate with the level of supercoiling *in vivo* and that the amount of Z-DNA formation was increased in a topoisomerase I mutant (372). Additionally, clusters of binding sites for topoisomerase II have been found in the Human Immunodeficiency Virus (HIV) integration region, which contain Z-DNA motifs, and cleavage activity of the enzyme was shown to occur at a Z-DNA-forming

region in the human beta-globin gene, which is a known hotspot for recombination and mutation *in vivo*, and the cleavage activity was found to be proportional to the length of the repeat (373). Thus, it is possible that topoisomerases participate in the “repair” mechanism of non-B DNA structures, particularly Z-DNA, causing DSBs and genomic instability related to translocations and human disease. We will first use our genomic yeast-based assay described in Chapters 2 and 3, to screen yeast strains deficient in topoisomerases I and II for an effect on H-DNA and Z-DNA-induced mutagenesis. We will follow up these studies in human cells using transient knockdown of these enzymes and H-DNA and Z-DNA-forming plasmids. Additionally, we could use viably active purified topoisomerases, to perform *in vitro* cleavage and/or binding assays of these enzymes at the non-B DNA-forming sites.

5.3. Ongoing Studies

5.3.1. Effect of modifications to non-B DNA structures on genomic instability.

H-DNA-forming and Z-DNA-forming sequences can comprise of guanine and/or cytosine-rich repeats and several of our (and others) studies implicate that the structure, rather than the sequence is mutagenic. Thus, it is of interest to determine if alterations made to the sequence and/or structure have an impact on the mutagenic capacity of non-B DNA structures. For example, Z-DNA (374-376), and triplex (e.g. H-DNA) (377,378) formation can be stabilized by methylation of cytosine residues at the C^5 position, which is very common in the human genome, particularly at CpG repeats in promoter regions (23). Therefore, this modification may provide further information on the influence of non-B DNA structures on gene expression. Preliminary data from our laboratory have shown that fully methylated Z-DNA-forming, CG repeats increase mutation frequency, replication fork stalling, and alterations in chromatin structure at the methylated Z-DNA regions (which is refractory to nucleosome assembly) in human cells (unpublished data), further suggesting that this modification can stabilize Z-DNA and enhance its mutagenic capabilities.

Additionally, alteration of guanine residues via oxidative damage, resulting from reactive oxygen species, can lead to strand breaks and genomic instability

(379). H-DNA-forming and Z-DNA-forming sequences are often guanine-rich, and thus may accumulate high levels of damage in this form at these sites within the cell, and it is of interest to determine if this base alteration effects non-B DNA structures *in vivo*. In theory, certain guanine residues within the H-DNA conformation may be protected against oxidative damage, as bases are stacked within the triple helix, while those guanine residues in the single-stranded loop may be more susceptible to damage as they are less protected than in a duplex (20,28,29). Moreover, the bases in Z-DNA reside facing the outside of the helix due to the *syn* conformation of the guanine residues and are less protected than bases in B-DNA, which are focused in the center of the helix (8,25). Thus we will determine if H-DNA-forming and Z-DNA-forming sequences are more susceptible to DNA damage induced by hydroxyl radicals and how these lesions affect non-B DNA-induced mutagenesis in mammalian cells. Preliminary data has shown an increase in mutation frequency with H-DNA that has been treated with hydroxyl radicals, suggesting an additive effect to the pre-determined mutagenic potential of H-DNA (125), whereby a damaged H-DNA-forming sequence further stimulates cellular repair processes to excise the damaged sequence, or damage within the non-B region is more resistant to repair (i.e., glycosylases that are responsible for removing bases damaged by oxidation, have diminished access for repair), leading to accumulation of damage. We will continue these studies using Z-DNA-forming sequences as well.

5.3.2. Determine the effects of non-B DNA structures on metabolism, nutrition, and aging, using an in vivo mouse model.

Obesity has become an alarming epidemic in the United States (and world-wide). It is estimated that 1/3 of adults and nearly 15% of children age 2-9 years have a body mass index (BMI) of >30 kg/m^2, placing them in the obese category (380). It is particularly alarming due to the connection of obesity to several chronic illnesses, such as cancers, including leukemias and lymphomas (381). Obesity has been shown to increase the risk cancer development and mortality rate (382-384), with nearly 20% of all cancer deaths in the United States attributed to overweight/obese body types (385,386). Consequently, understanding the relationship between obesity and cancer is an important step in cancer prevention. Interestingly, H-DNA-forming and Z-DNA-forming sequences have also been found at mutational "hotspots" related to leukemias and lymphomas (46,132,142,147,160,161,163,183,184,186-189,387). Additionally, mitochondrial DNA, which plays a crucial role in energy metabolism, has been shown to contain several non-B DNA-forming sequences, which have been implicated in mitochondrial deletion breakpoints [(388) and references therein]. Furthermore, diet has been shown to alter activity of the NHEJ repair pathway (389,390), and our previous results have implicated NHEJ in non-B DNA-induced mutagenesis in human cells (171,254). Therefore, there are several links between non-B DNA structures, obesity, and cancer, and we aim to

elucidate the underlying mechanisms contributing to these links.

Our laboratory, in collaboration with the laboratory of Dr. John DiGiovanni (Professor in the Division of Pharmacology and Toxicology at UT Austin), has developed a non-B DNA mouse model with stably integrated H-DNA-forming and Z-DNA-forming sequences upstream of a mutation-reporter gene integrated on mouse chromosomes, and shown that these integrations are mutagenic (128). We will use this mouse model to determine the effect of obesity on the mutagenic potential of non-B DNA, by feeding the mice high-fat (obese), normal, and calorie-restricted diets, and determining the non-B DNA-induced mutagenesis as previously described (128). We will also determine tissue-specific effects by examining various tissues in the mice, following treatment of the various diets. In this manner, the effects of obesity on genomic instability can be determined *in vivo*. Our hypothesis is that a high-fat diet, consistent with obesity and increased reactive oxygen species, will increase non-B DNA-induced genomic instability, while a calorie-restricted diet, may be protective against genomic instability. These studies are currently being conducted, and once completed will provide insights into the link between obesity and cancer etiology.

The incidence of cancer increases in an age-related fashion (391-393); as the aging process occurs, mutations (genetic instability) accumulate, increasing the prevalence for disease (394-398). The number of mutations in tumor cells has even been shown to dramatically increase with age (396). Increased age has also been shown to be associated with less efficient DNA repair (399-406).

Therefore, there is also evidence to suggest a link between genetic instability and age as a risk factor for cancer. While we have shown that H-DNA-forming and Z-DNA-forming sequences were mutagenic in the transgenic mouse model described above (128), these studies were performed in young mice (2-month old), and age was not a factor in our results. We will repeat the mutation-reporter mutagenesis studies using our mouse model, which contains integrated non-B DNA-forming sequences, as they age and determine what, if any, age has an effect on the level of non-B DNA-related genomic instability, and use PCR to identify sites of deletions and breaks in and around the non-B DNA-forming region. We will also determine if the effects are tissue dependent, as we will examine mutagenesis in various tissue types. Other studies have demonstrated that genetic instabilities related to hairpins, which form at triplet repeats and are associated with neurological disorders, have both an age and tissue-specific component (407,408). Thus, we predict that mutations will increase with age as a result of accumulation of damage to the DNA, and it will be interesting to determine the effect on non-B DNA-induced mutagenesis. These studies will provide insight into the mechanistic role of aging in cancer-associated genomic instability.

5.4. Long-term Goals

5.4.1. Continue to investigate non-B DNA-induced mutagenesis in a chromosomal context and using an in vivo mouse model.

While the studies presented here in Chapters 2 and 3 provide convincing evidence for the requirement of repair proteins in the mechanism responsible for non-B DNA-induced genomic instability, they are largely based on results obtained using plasmid mutation-reporter assays in human cells. However, to more fully mimic the mechanism of non-B DNA-stimulated mutagenesis in the context of chromatin, we will stably transfect Z-DNA-forming and H-DNA-forming sequences into human cell lines used in the studies described in Chapters 2 and 3, including wild-type and repair-deficient cell lines (i.e., XPF- and MSH2-deficient cells lines with Z-DNA-forming sequences). Regions containing non-B DNA inserts will be recovered as a mutation-reporter shuttle vector and subjected to blue/white screening as previously described (262) to determine mutation frequencies and spectra. Our *hypothesis* is that results from the chromosomally integrated non-B DNA-forming sequences will be similar to our plasmid-based assays, as we have already shown that these sequences are mutagenic in a chromosomal context in an *in vivo* mouse model (128), and on chromosomes in human cells (129).

Furthermore, we will continue our *in vivo* studies using our novel H-DNA and Z-DNA mutation-reporter mice (128) and will cross them with NER- or MMR-deficient mice to confirm our findings in an *in vivo*, mammalian model. These studies will be particularly useful in determining tissue-specific effects of non-B DNA mutagenesis, as we will test various tissues from the mouse model for levels of genomic instability related to Z-DNA and H-DNA in wild-type and repair-deficient mice. For example, as Z-DNA-forming and H-DNA-forming sequences map to breakpoints in leukemias and lymphomas, we can examine lymphocytes to determine if deletions/DSBs/translocations are more prevalent in these cells versus others. These studies will provide valuable information as to the involvement of repair proteins in non-B DNA-induced genomic instability in a physiologically relevant context.

5.4.2. Non-B DNA and disease-related therapeutics

The studies presented in Chapters 2 and 3 along with our previous studies (see Table 1.1) have clearly demonstrated that sequences with the capacity to form Z-DNA and H-DNA (and other non-B DNA structures) are mutagenic in multiple species, and that these sequences map to translocation breakpoint hotspots related to human disease, including cancers. We have begun to elucidate the mechanisms involved at the molecular level, and demonstrated that repair proteins from several pathways are involved. However, the importance of

acquiring such knowledge in the field of genomic instability is to ultimately design potential therapies to prevent diseases related to genetic instability, and/or to use our knowledge of how non-B DNA structures induce mutations to invoke genetic instability in cancer cells for *targeted* oncogenic therapeutics, that avoid the side effects of conventional chemotherapy on healthy cells. For example, the identification of non-B DNA *'destabilizing'* compounds (as outlined in section 5.2.5.), or specific domains of non-B DNA-binding repair proteins, could be used to design mutant repair proteins, in which domains required for repair function remain active, while the domain required to recognize non-B DNA structures is defective, preventing error-prone processing of such structures, thus decreasing non-B DNA-induced mutagenesis in normal cells.

Alternatively, identification of non-B DNA *'stabilizing'* compounds may be used to increase genomic instability and apoptosis in cancer cells using targeting gene therapy. As non-B DNA-forming sequences map to promoter regions, stabilizing compounds could be used to increase gene expression of tumor suppressor genes, such as p53 [one of the most prevalent abnormalities in human cancers (409-411)], by stabilizing Z-DNA (for example) in promoter regions to increase cell death in cancerous tumors. For example, if the human Z-DNA-binding protein, ADAR1, is found to have an effect on Z-DNA-induced mutagenesis in the studies proposed here (in section 5.2.1. and Chapter 4), targeted overexpression of this protein in cancer cells will increase instability in the cancer genome, potentially resulting in cell death.

Evidence has linked Z-DNA regulation of gene expression to cancer. For example, ADAM-12 is a multifunctional protein that is over-expressed in many cancers, and found to be required for progressive tumor progression and metastasis (412-420). MeCP2 is required for ADAM-12 suppression by interacting with a Z-DNA-forming negative regulatory element (NRE) in the promoter region of ADAM-12, and deficiency of MeCP2 in breast cancer cells resulted in overexpression of ADAM-12 leading to cell proliferation and metastasis (421). Identifying a compound that mimics MeCP2, by binding at the Z-DNA-related NRE, could serve as a potential targeted therapy for this type of breast cancer. Similar studies have been performed in various models relating to another non-B DNA structure, G4 DNA, which is found to be enriched at telomeres. In cancer cells, telomeres are maintained at indefinite length, and thus pose a potential therapeutic target. The G4 DNA structure and potential therapeutic mechanisms to modulate gene expression (including telomerase, *c-MYC,* vascular endothelial growth factor *(VEGF)*, and other genes implicated in cancer etiology) as a cancer therapy has been demonstrated in several studies [(217,422-425) and extensively reviewed in (426)], establishing great promise and potential for non-B DNA structures as an avenue for direct and more targeted cancer therapies.

5.5. Summary and Significance

Since the discovery of the canonical B-DNA double-helix more than 60 years ago by Watson and Crick (1), decades of research has been devoted to studying the structure of DNA, and several alternative DNA structures, collectively termed “non-B DNA”, have been identified and described (10-12). Non-B DNA structures, including Z-DNA (8,18) and H-DNA (19-21) (Fig. 1.2), do not conform to the Watson-Crick right-handed B-DNA, but rather result in significant alterations to the B-DNA double helix. For example, H-DNA is an intra-molecular triple helix structure in which three strands of DNA interact (28,29), and Z-DNA is a left-handed, double helical structure (8,25). Non-B DNA structures form at specific repetitive sequences, which are highly abundant in the human genome, and the distribution of non-B DNA-forming sequences co-localizes with promoter regions (22,23,40,80,81). H-DNA and Z-DNA have been implicated in the regulation of several DNA metabolic processes, including gene expression, and replication (41,42,49,50,82-86). Alternatively, H-DNA-forming and Z-DNA-forming sequences also map to regions prone to genomic instability related to human diseases, including genes involved in leukemias and lymphomas (84,147,160,177,179,186,207,269,285,289,378,387,424).

Prior to obtaining the results presented in this document, our laboratory demonstrated that Z-DNA-forming and H-DNA-forming sequences are mutagenic and can stimulate the formation of DSBs in various model systems (from bacteria

to human), and that these sequences map to translocation and deletion hotspots in human cancer genomes (Figs. 2.3 and 3.2), implicating non-B DNA structures in cancer etiology (Table 1.1 and references therein). However, the underlying mechanism(s) for Z-DNA and H-DNA-induced mutagenesis were unknown, and the goal of the studies presented in this document were to elucidate such mechanisms. We speculated that distortions induced in the canonical B-DNA helix by the formation of H-DNA and Z-DNA may be recognized as "DNA damage" by DNA repair recognition proteins, and evidence from our laboratory suggest that non-B DNA may serve as substrates for nucleotide excision repair (NER) and/or mismatch repair (MMR) pathways (120,126,130,231,252,253,324,427). Taken together, the evidence implicating NER and MMR proteins in non-B DNA-induced mutagenesis, we **hypothesized that the helical distortions induced by non-B DNA are recognized as "damage" and processed by DNA repair machinery via an error-generating mechanism resulting in genomic instability.** We proposed three specific aims to address this hypothesis:

Specific Aim 1 - Determine the extent to which NER and MMR proteins impact H-DNA structure-induced mutagenesis in eukaryotes.

Specific Aim 2 - Determine the extent to which NER and MMR proteins impact Z-DNA structure-induced mutagenesis in eukaryotes.

Specific Aim 3 - Characterize the function/pathway or the proteins involved in non-B-DNA-induced mutagenesis.

In Specific Aims 1 and 3 (addressed in Chapter 3), we investigated the mechanism(s) responsible for genomic instability caused by H-DNA. Using a yeast artificial chromosome (YAC), we demonstrated *for the first time* that a naturally occurring, H-DNA-forming sequence is mutagenic in wild-type *Saccharomyces cerevisiae* cells when compared to a control sequence, and that H-DNA induced DSBs on the YAC (Fig. 3.3b and c), consistent with our findings of H-DNA-induced DSBs and genetic instability in mammalian cells and mice (125,128). Using a yeast deletion library, we screened repair-deficient cells and found that H-DNA-induced mutagenesis was suppressed in yeast cells deficient in Rad10(ERCC1) and Rad1(XPF) (Fig. 3.3b).

Using plasmid-based mutation-reporter assays in human cells, we confirmed our results in yeast, demonstrating that H-DNA-induced mutagenesis was suppressed in human cells deficient in the NER nucleases, ERCC1-XPF and XPG; and in contrast, H-DNA-induced genetic instability was stimulated in cells deficient in FEN1, a replication-related endonuclease (Figs. 3.4 and 3.7; and Table 3.2), providing evidence for replication-dependent and replication-independent models of H-DNA-induced genetic instability in eukaryotes. Using ligation-mediated PCR, we demonstrated that DSBs were formed near the H-DNA-forming site in wild-type human cells, which were absent in human XPF-deficient cells (Fig. 3.5a). Using an H-DNA-forming oligonucleotide, we identified the structure as a novel substrate for cleavage by ERCC1-XPF, XPG, and FEN1

(Fig. 3.5 and 3.7) and determined that these proteins function in distinct replication-independent and replication-dependent pathways of genetic instability (Fig. 3.9). Thus, we propose a model for H-DNA-induced mutagenesis whereby NER proteins recognize and process H-DNA, leading to DSBs in the absence of replication, while FEN1 cleavage of H-DNA can result in error-free processing during replication. However, in the absence of FEN1, the H-DNA structure may persist at a replication fork, causing fork collapse, which could result in DSBs and subsequent deletions (Fig. 3.10).

In Specific Aims 2 and 3 (addressed in Chapter 2), we investigated the mechanism(s) responsible for Z-DNA-induced mutagenesis in yeast and human cells. Using similar assays to the H-DNA studies, we used a YAC containing a Z-DNA-forming or B-DNA-forming control sequence (Fig. 2.6) to screen a yeast strains deficient in DNA repair proteins. We demonstrated *for the first time* that a Z-DNA-forming sequence is mutagenic in wild-type *S. cerevisiae* cells when compared to a control sequence (Fig. 2.7). Interestingly, we found that in yeast strains deficient in NER and MMR protein complexes [Rad10-Rad1 (XPF-ERCC1), and Msh2-Msh3, respectively], Z-DNA-induced mutagenesis was significantly decreased (Figs. 2.9 and 2.10; and Table 2.3), implicating these proteins in non-B DNA-induced genomic instability. Additionally, we demonstrated that in wild-type strains, DSBs in the Z-DNA-forming YAC were higher than in the B-DNA-forming control YAC, which were decreased in Rad10, Rad1, Msh2, or Msh3-deficient strains (Fig. 2.13).

Again, we confirmed our yeast results using a plasmid-based mutation-reporter assay in wild-type and repair-deficient human cells. The results in human cells validated the yeast data; the increase in mutations due to the Z-DNA-forming plasmid in wild-type cells was abrogated in cells deficient in XPF or MSH2 (Fig. 2.18). Using chromatin immunoprecipitation (ChIP) assays, we demonstrated that ERCC1-XPF and MSH2-MSH3 complexes were significantly enriched at Z-DNA-forming sequences relative to B-DNA-forming sequences, and that ERCC1-XPF recruitment was dependent on MSH2-MSH3 (Fig. 2.19). Furthermore, using an *in vitro* primer extension assay, we found that ERCC1-XPF cleaved DNA near the Z-DNA-forming site in human whole cell extracts (Fig. 2.22). We propose a novel relationship in which these NER and MMR complexes (acting outside of their canonical repair functions) recognize and attempt to repair Z-DNA in eukaryotic genomes, potentially resulting in error-prone processing causing DSBs (Fig. 2.23). Thus, our results implicate Z-DNA in cancer etiology.

The results obtained with Z-DNA are in contrast to the H-DNA studies, which suggested that proteins from a single pathway, NER, are involved in H-DNA-induced mutagenesis, rather than certain components from both NER and MMR, as seems to be the case for Z-DNA-induced mutagenesis. This is not particularly surprising, as the structures of H-DNA and Z-DNA are inherently different, thus it is likely that different mechanisms may be required for mutagenesis caused by the two structures. Furthermore, the H-DNA structure (Fig. 3.1b) more closely resembles bulky adducts that are recognized by NER

factors, while the B-Z junctions in Z-DNA (Fig. 2.1b) resemble the small loops recognized by MMR factors, yet the significant helical distortions caused by the left-handed conformation of Z-DNA may stimulate processing by NER factors, and thus it is conceivable that factors from both mechanisms are involved in Z-DNA recognition and processing.

Several significant findings were established in the studies presented here. First, we have shown for the *first time* that H-DNA and Z-DNA are mutagenic in a new model organism, *S. cerevisiae*, further implicating the evolutionary importance of non-B DNA structures, as they function across multiple species. We have presented evidence to support an intra-molecular H-DNA structure as a novel substrate for the ERCC1-XPF, XPF, and FEN1 structure-specific nucleases, and that ERCC1-XPF can cleave near a Z-DNA-forming sequence. We have identified specific components of DNA repair pathways to function in non-B DNA-induced mechanisms in distinct, structure-specific pathways, which present novel substrates for NER and MMR proteins. We propose the first descriptions of non-B DNA-induced mutagenesis that are consistent across multiple species of eukaryotes, including a unique relationship and crosstalk between DNA repair complexes from the NER and MMR pathways required for Z-DNA-induced mutagenesis, and replication-independent and replication-dependent pathways involved in H-DNA-induced mutagenesis. These mechanisms provide insight to an error-generating mechanism by which non-B DNA structures are recognized as “damage” by various repair factors, and are

processed in a manner that potentially leads to DSBs, which may be responsible for naturally-occurring, endogenous mutation “hotspots” in human cancer.

This work clearly advances our understanding of the physiologically relevant connection of non-B DNA-forming sequences with translocation “hotspots” in human cancers such as Burkitt’s lymphoma and leukemias. Our findings implicate non-B DNA-forming sequences as an endogenous contributor to human disease and cancer etiology, and a better understanding of the mechanisms involved will allow for advancement in targeted therapeutics to prevent and/or treat diseases of genetic instability.

References

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
2. Franklin, R.E. and Gosling, R.G. (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, **172**, 156-157.
3. Franklin, R.E. and Gosling, R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740-741.
4. Wilkins, M.H., Seeds, W.E., Stokes, A.R. and Wilson, H.R. (1953) Helical structure of crystalline deoxypentose nucleic acid. *Nature*, **172**, 759-762.
5. Wilkins, M.H., Stokes, A.R. and Wilson, H.R. (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738-740.
6. Crick, F.H. (1954) The Complementary Structure of DNA. *Proc Natl Acad Sci U S A*, **40**, 756-758.
7. Karp, G. (2008) *Cell and molecular biology : concepts and experiments*. 5th ed. John Wiley, Chichester.
8. Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680-686.
9. Pierce, B.A. (2005) *Genetics : a conceptual approach*. 2nd ed. W.H. Freeman, New York.
10. Wells, R.D. (1988) Unusual DNA structures. *J Biol Chem*, **263**, 1095-1098.
11. Wang, G. and Vasquez, K.M. (2006) Non-B DNA structure-induced genetic instability. *Mutat Res*, **598**, 103-119.
12. Choi, J. and Majima, T. (2011) Conformational changes of non-B DNA. *Chem Soc Rev*, **40**, 5893-5909.
13. Lilley, D.M. (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci U S A*, **77**, 6468-6472.
14. Panayotatos, N. and Wells, R.D. (1981) Cruciform structures in supercoiled DNA. *Nature*, **289**, 466-470.
15. Lyamichev, V.I., Panyutin, I.G. and Frank-Kamenetskii, M.D. (1983) Evidence of cruciform structures in superhelical DNA provided by two-dimensional gel electrophoresis. *FEBS Lett*, **153**, 298-302.
16. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364-366.
17. Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*, **344**, 410-414.
18. Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G. and Kim, K.K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, **437**, 1183-1186.

19. Htun, H. and Dahlberg, J.E. (1988) Single strands, triple strands, and kinks in H-DNA. *Science*, **241**, 1791-1796.
20. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1986) Structures of homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn*, **3**, 667-669.
21. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1987) DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature*, **330**, 495-497.
22. Sinden, R.R. (1994) *DNA structure and function*. Academic Press, San Diego.
23. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
24. Zhao, J., Bacolla, A., Wang, G. and Vasquez, K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci*, **67**, 43-62.
25. Herbert, A. and Rich, A. (1999) Left-handed Z-DNA: structure and function. *Genetica*, **106**, 37-47.
26. Singleton, C.K., Klysik, J., Stirdivant, S.M. and Wells, R.D. (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature*, **299**, 312-316.
27. Nordheim, A., Lafer, E.M., Peck, L.J., Wang, J.C., Stollar, B.D. and Rich, A. (1982) Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding. *Cell*, **31**, 309-318.
28. Mirkin, S.M. and Frank-Kamenetskii, M.D. (1994) H-DNA and related structures. *Annu Rev Biophys Biomol Struct*, **23**, 541-576.
29. Htun, H. and Dahlberg, J.E. (1989) Topology and formation of triple-stranded H-DNA. *Science*, **243**, 1571-1576.
30. Pilch, D.S., Levenson, C. and Shafer, R.H. (1991) Structure, stability, and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. *Biochemistry*, **30**, 6081-6088.
31. Jain, A., Wang, G. and Vasquez, K.M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, **90**, 1117-1130.
32. Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *The Journal of biological chemistry*, **282**, 32433-32441.
33. Sinden, R.R. and Pettijohn, D.E. (1984) Cruciform transitions in DNA. *J Biol Chem*, **259**, 6593-6600.
34. Smith, G.R. (2008) Meeting DNA palindromes head-to-head. *Genes Dev*, **22**, 2612-2620.
35. Oussatcheva, E.A., Pavlicek, J., Sankey, O.F., Sinden, R.R., Lyubchenko, Y.L. and Potaman, V.N. (2004) Influence of global DNA topology on cruciform formation in supercoiled DNA. *J Mol Biol*, **338**, 735-743.

36. Watson, J., Hays, F.A. and Ho, P.S. (2004) Definitions and analysis of DNA Holliday junction geometry. *Nucleic Acids Res*, **32**, 3017-3027.
37. Nag, D.K. and Petes, T.D. (1991) Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in Saccharomyces cerevisiae. *Genetics*, **129**, 669-673.
38. Losch, F.O., Bredenbeck, A., Hollstein, V.M., Walden, P. and Wrede, P. (2007) Evidence for a large double-cruciform DNA structure on the X chromosome of human and chimpanzee. *Hum Genet*, **122**, 337-343.
39. Majumdar, A. and Patel, D.J. (2002) Identifying hydrogen bond alignments in multistranded DNA architectures by NMR. *Acc Chem Res*, **35**, 1-11.
40. Cox, R. and Mirkin, S.M. (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci U S A*, **94**, 5237-5242.
41. Khuu, P., Sandor, M., DeYoung, J. and Ho, P.S. (2007) Phylogenomic analysis of the emergence of GC-rich transcription elements. *P Natl Acad Sci USA*, **104**, 16528-16533.
42. Schroth, G.P. and Ho, P.S. (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res*, **23**, 1977-1983.
43. Nordheim, A., Pardue, M.L., Lafer, E.M., Moller, A., Stollar, B.D. and Rich, A. (1981) Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes. *Nature*, **294**, 417-422.
44. Lancillotti, F., Lopez, M.C., Arias, P. and Alonso, C. (1987) Z-DNA in transcriptionally active chromosomes. *Proc Natl Acad Sci U S A*, **84**, 1560-1564.
45. Lipps, H.J., Nordheim, A., Lafer, E.M., Ammermann, D., Stollar, B.D. and Rich, A. (1983) Antibodies against Z DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate stylonychia mytilus. *Cell*, **32**, 435-441.
46. Wolfl, S., Wittig, B. and Rich, A. (1995) Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim Biophys Acta*, **1264**, 294-302.
47. Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W. and Rich, A. (1992) Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J*, **11**, 4653-4663.
48. Nordheim, A. and Rich, A. (1983) Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. *Nature*, **303**, 674-679.
49. Wolfl, S., Martinez, C., Rich, A. and Majzoub, J.A. (1996) Transcription of the human corticotropin-releasing hormone gene in NPLC cells is correlated with Z-DNA formation. *Proc Natl Acad Sci U S A*, **93**, 3664-3668.

50. Wolfl, S., Wittig, B., Dorbic, T. and Rich, A. (1997) Identification of processes that influence negative supercoiling in the human c-myc gene. *Biochim Biophys Acta*, **1352**, 213-221.
51. Santoro, C., Costanzo, F. and Ciliberto, G. (1984) Inhibition of eukaryotic tRNA transcription by potential Z-DNA sequences. *EMBO J*, **3**, 1553-1559.
52. Peck, L.J. and Wang, J.C. (1985) Transcriptional block caused by a negative supercoiling induced structural change in an alternating CG sequence. *Cell*, **40**, 129-137.
53. de Mercoyrol, L., Corda, Y., Job, C. and Job, D. (1992) Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA. *Eur J Biochem*, **206**, 49-58.
54. Sheridan, S.D., Benham, C.J. and Hatfield, G.W. (1999) Inhibition of DNA supercoiling-dependent transcriptional activation by a distant B-DNA to Z-DNA transition. *J Biol Chem*, **274**, 8169-8174.
55. Liu, R., Liu, H., Chen, X., Kirby, M., Brown, P.O. and Zhao, K. (2001) Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell*, **106**, 309-318.
56. Rothenburg, S., Koch-Nolte, F., Rich, A. and Haag, F. (2001) A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity. *Proc Natl Acad Sci U S A*, **98**, 8985-8990.
57. Sheridan, S.D., Opel, M.L. and Hatfield, G.W. (2001) Activation and repression of transcription initiation by a distant DNA structural transition. *Mol Microbiol*, **40**, 684-690.
58. Liu, H., Mulholland, N., Fu, H. and Zhao, K. (2006) Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling. *Mol Cell Biol*, **26**, 2550-2559.
59. Wong, B., Chen, S., Kwon, J.A. and Rich, A. (2007) Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **104**, 2229-2234.
60. Ditlevson, J.V., Tornaletti, S., Belotserkovskii, B.P., Teijeiro, V., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2008) Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res*, **36**, 3163-3170.
61. Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U. and Rich, A. (2001) Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat Struct Biol*, **8**, 761-765.
62. Liu, Y., Wolff, K.C., Jacobs, B.L. and Samuel, C.E. (2001) Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity. *Virology*, **289**, 378-387.
63. Kahmann, J.D., Wecking, D.A., Putter, V., Lowenhaupt, K., Kim, Y.G., Schmieder, P., Oschkinat, H., Rich, A. and Schade, M. (2004) The solution structure of the N-terminal domain of E3L shows a tyrosine

conformation that may explain its reduced affinity to Z-DNA in vitro. *Proc Natl Acad Sci U S A*, **101**, 2712-2717.
64. Kwon, J.A. and Rich, A. (2005) Biological function of the vaccinia virus Z-DNA-binding protein E3L: gene transactivation and antiapoptotic activity in HeLa cells. *Proc Natl Acad Sci U S A*, **102**, 12759-12764.
65. Kim, Y.G., Lowenhaupt, K., Oh, D.B., Kim, K.K. and Rich, A. (2004) Evidence that vaccinia virulence factor E3L binds to Z-DNA in vivo: Implications for development of a therapy for poxvirus infection. *Proc Natl Acad Sci U S A*, **101**, 1514-1518.
66. Lafer, E.M., Sousa, R.J. and Rich, A. (1988) Z-DNA-binding proteins in Escherichia coli purification, generation of monoclonal antibodies and gene isolation. *J Mol Biol*, **203**, 511-516.
67. Krishna, P., Kennedy, B.P., Waisman, D.M., van de Sande, J.H. and McGhee, J.D. (1990) Are many Z-DNA binding proteins actually phospholipid-binding proteins? *Proc Natl Acad Sci U S A*, **87**, 1292-1295.
68. Zhang, S., Lockshin, C., Herbert, A., Winter, E. and Rich, A. (1992) Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae. *EMBO J*, **11**, 3787-3796.
69. Nordheim, A., Tesser, P., Azorin, F., Kwon, Y.H., Moller, A. and Rich, A. (1982) Isolation of Drosophila proteins that bind selectively to left-handed Z-DNA. *Proc Natl Acad Sci U S A*, **79**, 7729-7733.
70. Arndt-Jovin, D.J., Udvardy, A., Garner, M.M., Ritter, S. and Jovin, T.M. (1993) Z-DNA binding and inhibition by GTP of Drosophila topoisomerase II. *Biochemistry*, **32**, 4862-4872.
71. Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K. and Rich, A. (2005) A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc Natl Acad Sci U S A*, **102**, 1602-1607.
72. Rohner, K.J., Hobi, R. and Kuenzle, C.C. (1990) Z-DNA-binding proteins. Identification critically depends on the proper choice of ligands. *J Biol Chem*, **265**, 19112-19115.
73. Herbert, A.G., Spitzner, J.R., Lowenhaupt, K. and Rich, A. (1993) Z-DNA binding protein from chicken blood nuclei. *Proc Natl Acad Sci U S A*, **90**, 3339-3342.
74. Rothenburg, S., Schwartz, T., Koch-Nolte, F. and Haag, F. (2002) Complex regulation of the human gene for the Z-DNA binding protein DLM-1. *Nucleic Acids Res*, **30**, 993-1000.
75. Leith, I.R., Hay, R.T. and Russell, W.C. (1988) Detection of Z DNA binding proteins in tissue culture cells. *Nucleic Acids Res*, **16**, 8277-8289.
76. Pham, H.T., Park, M.Y., Kim, K.K., Kim, Y.G. and Ahn, J.H. (2006) Intracellular localization of human ZBP1: Differential regulation by the Z-DNA binding domain, Zalpha, in splice variants. *Biochem Biophys Res Commun*, **348**, 145-152.

77. Wang, G. and Vasquez, K.M. (2007) Z-DNA, an active element in the genome. *Front Biosci*, **12**, 4424-4438.
78. Rich, A. and Zhang, S. (2003) Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet*, **4**, 566-572.
79. Felsenfeld, G. and Rich, A. (1957) Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta*, **26**, 457-468.
80. Behe, M.J. (1995) An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Res*, **23**, 689-695.
81. Bucher, P. and Yagil, G. (1991) Occurrence of oligopurine.oligopyrimidine tracts in eukaryotic and prokaryotic genes. *DNA Seq*, **1**, 157-172.
82. Bacolla, A., Collins, J.R., Gold, B., Chuzhanova, N., Yi, M., Stephens, R.M., Stefanov, S., Olsh, A., Jakupciak, J.P., Dean, M. *et al.* (2006) Long homopurine*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res*, **34**, 2663-2675.
83. Praseuth, D., Guieysse, A.L. and Helene, C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim Biophys Acta*, **1489**, 181-206.
84. Kato, M. and Shimizu, N. (1992) Effect of the potential triplex DNA region on the in vitro expression of bacterial beta-lactamase gene in superhelical recombinant plasmids. *J Biochem*, **112**, 492-494.
85. Sarkar, P.S. and Brahmachari, S.K. (1992) Intramolecular triplex potential sequence within a gene down regulates its expression in vivo. *Nucleic Acids Res*, **20**, 5713-5718.
86. Michel, D., Chatelain, G., Herault, Y., Harper, F. and Brun, G. (1993) H-DNA can act as a transcriptional insulator. *Cell Mol Biol Res*, **39**, 131-140.
87. Firulli, A.B., Maibenco, D.C. and Kinniburgh, A.J. (1994) Triplex forming ability of a c-myc promoter element predicts promoter strength. *Arch Biochem Biophys*, **310**, 236-242.
88. Duval-Valentin, G., de Bizemont, T., Takasugi, M., Mergny, J.L., Bisagni, E. and Helene, C. (1995) Triple-helix specific ligands stabilize H-DNA conformation. *J Mol Biol*, **247**, 847-858.
89. Grabczyk, E. and Fishman, M.C. (1995) A long purine-pyrimidine homopolymer acts as a transcriptional diode. *J Biol Chem*, **270**, 1791-1797.
90. Xu, G. and Goodridge, A.G. (1996) Characterization of a polypyrimidine/polypurine tract in the promoter of the gene for chicken malic enzyme. *J Biol Chem*, **271**, 16008-16019.
91. Brahmachari, S.K., Sarkar, P.S., Raghavan, S., Narayan, M. and Maiti, A.K. (1997) Polypurine/polypyrimidine sequences as cis-acting transcriptional regulators. *Gene*, **190**, 17-26.

92. Maiti, A.K. and Brahmachari, S.K. (2001) Poly purine.pyrimidine sequences upstream of the beta-galactosidase gene affect gene expression in Saccharomyces cerevisiae. *BMC Mol Biol*, **2**, 11.
93. Rustighi, A., Tessari, M.A., Vascotto, F., Sgarra, R., Giancotti, V. and Manfioletti, G. (2002) A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. *Biochemistry*, **41**, 1229-1240.
94. Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M. and Hanawalt, P.C. (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem*, **282**, 32433-32441.
95. Grabczyk, E. and Usdin, K. (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res*, **28**, 2815-2822.
96. Grabczyk, E. and Usdin, K. (2000) Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. *Nucleic Acids Res*, **28**, 4930-4937.
97. Pahwa, G.S., Maher, L.J., 3rd and Hollingsworth, M.A. (1996) A potential H-DNA element in the MUC1 promoter does not influence transcription. *J Biol Chem*, **271**, 26543-26546.
98. Lu, Q., Teare, J.M., Granok, H., Swede, M.J., Xu, J. and Elgin, S.C. (2003) The capacity to form H-DNA cannot substitute for GAGA factor binding to a (CT)n*(GA)n regulatory site. *Nucleic Acids Res*, **31**, 2483-2494.
99. Gilmour, D.S., Thomas, G.H. and Elgin, S.C. (1989) Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science*, **245**, 1487-1490.
100. O'Neill, D., Bornschlegel, K., Flamm, M., Castle, M. and Bank, A. (1991) A DNA-binding factor in adult hematopoietic cells interacts with a pyrimidine-rich domain upstream from the human delta-globin gene. *Proc Natl Acad Sci U S A*, **88**, 8953-8957.
101. Yee, H.A., Wong, A.K., van de Sande, J.H. and Rattner, J.B. (1991) Identification of novel single-stranded d(TC)n binding proteins in several mammalian species. *Nucleic Acids Res*, **19**, 949-953.
102. Muraiso, T., Nomoto, S., Yamazaki, H., Mishima, Y. and Kominami, R. (1992) A single-stranded DNA binding protein from mouse tumor cells specifically recognizes the C-rich strand of the (AGG:CCT)n repeats that can alter DNA conformation. *Nucleic Acids Res*, **20**, 6631-6635.
103. Kolluri, R. and Kinniburgh, A.J. (1991) Full length cDNA sequence encoding a nuclease-sensitive element DNA binding protein. *Nucleic Acids Res*, **19**, 4771.

104. Kolluri, R., Torrey, T.A. and Kinniburgh, A.J. (1992) A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucleic Acids Res*, **20**, 111-116.
105. Aharoni, A., Baran, N. and Manor, H. (1993) Characterization of a multisubunit human protein which selectively binds single stranded d(GA)n and d(GT)n sequence repeats in DNA. *Nucleic Acids Res*, **21**, 5221-5228.
106. Guieysse, A.L., Praseuth, D. and Helene, C. (1997) Identification of a triplex DNA-binding protein from human cells. *J Mol Biol*, **267**, 289-298.
107. Nelson, L.D., Musso, M. and Van Dyke, M.W. (2000) The yeast STM1 gene encodes a purine motif triple helical DNA-binding protein. *J Biol Chem*, **275**, 5573-5581.
108. Kim, Y.S. and Kang, H.S. (1989) Sequence-specific functions of the early palindrome domain within the SV40 core origin of replication. *Nucleic Acids Res*, **17**, 9279-9289.
109. Lin, S. and Kowalski, D. (1994) DNA helical instability facilitates initiation at the SV40 replication origin. *J Mol Biol*, **235**, 496-507.
110. Babendure, J.R., Babendure, J.L., Ding, J.H. and Tsien, R.Y. (2006) Control of mammalian translation by mRNA structure near caps. *RNA*, **12**, 851-861.
111. Repping, S., van Daalen, S.K., Brown, L.G., Korver, C.M., Lange, J., Marszalek, J.D., Pyntikova, T., van der Veen, F., Skaletsky, H., Page, D.C. *et al.* (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet*, **38**, 463-467.
112. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, **33**, 2908-2916.
113. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*, **33**, 2901-2907.
114. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.
115. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871-880.
116. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, **35**, 406-413.
117. Huppert, J.L., Bugaut, A., Kumari, S. and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res*, **36**, 6260-6268.
118. Godde, J.S., Kass, S.U., Hirst, M.C. and Wolffe, A.P. (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. *J Biol Chem*, **271**, 24325-24328.
119. Wong, H.M. and Huppert, J.L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol Biosyst*, **5**, 1713-1719.

120. Vasquez, K.M., Christensen, J., Li, L., Finch, R.A. and Glazer, P.M. (2002) Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. *P Natl Acad Sci USA*, **99**, 5848-5853.
121. Thoma, B.S. and Vasquez, K.M. (2003) Critical DNA damage recognition functions of XPC-hHR23B and XPA-RPA in nucleotide excision repair. *Mol Carcinog*, **38**, 1-13.
122. Bacolla, A., Jaworski, A., Larson, J.E., Jakupciak, J.P., Chuzhanova, N., Abeysinghe, S.S., O'Connell, C.D., Cooper, D.N. and Wells, R.D. (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. *P Natl Acad Sci USA*, **101**, 14162-14167.
123. Bacolla, A. and Wells, R.D. (2004) Non-B DNA conformations, genomic rearrangements, and human disease. *The Journal of biological chemistry*, **279**, 47411-47414.
124. Christensen, L.A., Conti, C.J., Fischer, S.M. and Vasquez, K.M. (2004) Mutation frequencies in murine keratinocytes as a function of carcinogenic status. *Mol Carcinog*, **40**, 122-133.
125. Wang, G. and Vasquez, K.M. (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *P Natl Acad Sci USA*, **101**, 13448-13453.
126. Thoma, B.S., Wakasugi, M., Christensen, J., Reddy, M.C. and Vasquez, K.M. (2005) Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. *Nucleic Acids Res*, **33**, 2993-3001.
127. Wang, G., Christensen, L.A. and Vasquez, K.M. (2006) Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *P Natl Acad Sci USA*, **103**, 2677-2682.
128. Wang, G., Carbajal, S., Vijg, J., DiGiovanni, J. and Vasquez, K.M. (2008) DNA structure-induced genomic instability in vivo. *J Natl Cancer Inst*, **100**, 1815-1817.
129. Betous, R., Rey, L., Wang, G., Pillaire, M.J., Puget, N., Selves, J., Biard, D.S., Shin-ya, K., Vasquez, K.M., Cazaux, C. *et al.* (2009) Role of TLS DNA polymerases eta and kappa in processing naturally occurring structured DNA in human cells. *Mol Carcinog*, **48**, 369-378.
130. Zhao, J., Jain, A., Iyer, R.R., Modrich, P.L. and Vasquez, K.M. (2009) Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res*, **37**, 4420-4429.
131. Jain, A., Bacolla, A., Chakraborty, P., Grosse, F. and Vasquez, K.M. (2010) Human DHX9 helicase unwinds triple-helical DNA structures. *Biochemistry*, **49**, 6992-6999.
132. Bacolla, A., Wang, G., Jain, A., Chuzhanova, N.A., Cer, R.Z., Collins, J.R., Cooper, D.N., Bohr, V.A. and Vasquez, K.M. (2011) Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of

base substitution in human cells. *The Journal of biological chemistry*, **286**, 10017-10026.
133. Lu, S., Wang, G., Bacolla, A., Zhao, J., Spitser, S. and Vasquez, K.M. (2015) Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep*.
134. McLean, M.J. and Wells, R.D. (1988) The role of DNA sequence in the formation of Z-DNA versus cruciforms in plasmids. *J Biol Chem*, **263**, 7370-7377.
135. Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M. and McMurray, C.T. (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature structural & molecular biology*, **12**, 663-670.
136. Petruska, J., Arnheim, N. and Goodman, M.F. (1996) Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res*, **24**, 1992-1998.
137. Bowater, R.P. and Wells, R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog Nucleic Acid Res Mol Biol*, **66**, 159-202.
138. Caskey, C.T., Pizzuti, A., Fu, Y.H., Fenwick, R.G., Jr. and Nelson, D.L. (1992) Triplet repeat mutations in human disease. *Science*, **256**, 784-789.
139. Wells, R.D. (1996) Molecular basis of genetic instability of triplet repeats. *J Biol Chem*, **271**, 2875-2878.
140. Pearson, C.E., Zorbas, H., Price, G.B. and Zannis-Hadjopoulos, M. (1996) Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J Cell Biochem*, **63**, 1-22.
141. Kinniburgh, A.J. (1989) A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. *Nucleic Acids Res*, **17**, 7771-7778.
142. Raghavan, S.C., Chastain, P., Lee, J.S., Hegde, B.G., Houston, S., Langen, R., Hsieh, C.L., Haworth, I.S. and Lieber, M.R. (2005) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *J Biol Chem*, **280**, 22749-22760.
143. Malfoy, B., Rousseau, N., Vogt, N., Viegas-Pequignot, E., Dutrillaux, B. and Leng, M. (1986) Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes. *Nucleic Acids Res*, **14**, 3197-3214.
144. Johnston, B.H. (1992) Generation and detection of Z-DNA. *Methods Enzymol*, **211**, 127-158.
145. Feigon, J., Wang, A.H., van der Marel, G.A., van Boom, J.H. and Rich, A. (1985) Z-DNA forms without an alternating purine-pyrimidine sequence in solution. *Science*, **230**, 82-84.

146. Eichman, B.F., Schroth, G.P., Basham, B.E. and Ho, P.S. (1999) The intrinsic structure and stability of out-of-alternation base pairs in Z-DNA. *Nucleic Acids Res*, **27**, 543-550.
147. Boehm, T., Mengle-Gaw, L., Kees, U.R., Spurr, N., Lavenir, I., Forster, A. and Rabbitts, T.H. (1989) Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. *EMBO J*, **8**, 2621-2631.
148. Collins, J., Volckaert, G. and Nevers, P. (1982) Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of recA-independent deletion events in Escherichia coli. *Gene*, **19**, 139-146.
149. Cromie, G.A., Millar, C.B., Schmidt, K.H. and Leach, D.R. (2000) Palindromes as substrates for multiple pathways of recombination in Escherichia coli. *Genetics*, **154**, 513-522.
150. Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E. and Resnick, M.A. (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol Cell Biol*, **13**, 5315-5322.
151. Murphy, K.E. and Stringer, J.R. (1986) RecA independent recombination of poly[d(GT)-d(CA)] in pBR322. *Nucleic Acids Res*, **14**, 7325-7340.
152. Blaho, J.A. and Wells, R.D. (1989) Left-handed Z-DNA and genetic recombination. *Prog Nucleic Acid Res Mol Biol*, **37**, 107-126.
153. Wahls, W.P., Wallace, L.J. and Moore, P.D. (1990) The Z-DNA motif d(TG)30 promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. *Mol Cell Biol*, **10**, 785-793.
154. Kobori, J.A., Strauss, E., Minard, K. and Hood, L. (1986) Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. *Science*, **234**, 173-179.
155. Wahls, W.P. (1998) Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change. *Curr Top Dev Biol*, **37**, 37-75.
156. Cer, R.Z., Bruce, K.H., Donohue, D.E., Temiz, N.A., Mudunuri, U.S., Yi, M., Volfovsky, N., Bacolla, A., Luke, B.T., Collins, J.R. *et al.* (2012) Searching for non-B DNA-forming motifs using nBMST (non-B DNA motif search tool). *Curr Protoc Hum Genet*, **Chapter 18**, Unit 18 17 11-22.
157. Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res*, **34**, W676-682.
158. Wang, G., Gaddis, S. and Vasquez, K.M. (2013) Methods to detect replication-dependent and replication-independent DNA structure-induced genetic instability. *Methods*, **64**, 67-72.
159. Yunis, J.J., Frizzera, G., Oken, M.M., McKenna, J., Theologides, A. and Arnesen, M. (1987) Multiple recurrent genomic defects in follicular lymphoma. A possible model for cancer. *N Engl J Med*, **316**, 79-84.

160. Adachi, M. and Tsujimoto, Y. (1990) Potential Z-DNA elements surround the breakpoints of chromosome translocation within the 5' flanking region of bcl-2 gene. *Oncogene*, **5**, 1653-1657.
161. Seite, P., Hillion, J., Leroux, D., Berger, R. and Larsen, C.J. (1993) Common sequence in chromosome translocations affecting B- and T-cell malignancies: a novel recombination site? *Genes Chromosomes Cancer*, **6**, 253-254.
162. Rabbitts, T.H. and Boehm, T. (1991) Structural and functional chimerism results from chromosomal translocation in lymphoid tumors. *Adv Immunol*, **50**, 119-146.
163. Rimokh, R., Rouault, J.P., Wahbi, K., Gadoux, M., Lafage, M., Archimbaud, E., Charrin, C., Gentilhomme, O., Germain, D., Samarut, J. *et al.* (1991) A chromosome 12 coding region is juxtaposed to the MYC protooncogene locus in a t(8;12)(q24;q22) translocation in a case of B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer*, **3**, 24-36.
164. Xia, Y., Brown, L., Tsan, J.T., Yang, C.Y., Siciliano, M.J., Crist, W.M., Carroll, A.J. and Baer, R. (1992) The translocation (1;14)(p34;q11) in human T-cell leukemia: chromosome breakage 25 kilobase pairs downstream of the TAL1 protooncogene. *Genes Chromosomes Cancer*, **4**, 211-216.
165. Sinclair, P.B., Parker, H., An, Q., Rand, V., Ensor, H., Harrison, C.J. and Strefford, J.C. (2011) Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Hum Mol Genet*, **20**, 2591-2602.
166. Bacolla, A., Tainer, J.A., Vasquez, K.M. and Cooper, D.N. (2016) Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res*.
167. Casasnovas, J.M., Ellison, M.J., Rodriguez-Campos, A. and Azorin, F. (1987) The obtention of simian virus 40 recombinants carrying d(CG.GC)n, d(CA.GT)n and d(CT.GA)n sequences. Stability of the inserted simple repeating sequences. *Eur J Biochem*, **167**, 489-492.
168. Freund, A.M., Bichara, M. and Fuchs, R.P. (1989) Z-DNA-forming sequences are spontaneous deletion hot spots. *Proc Natl Acad Sci U S A*, **86**, 7465-7469.
169. Sinden, R.R. (2001) Neurodegenerative diseases. Origins of instability. *Nature*, **411**, 757-758.
170. Wang, G., Christensen, L.A. and Vasquez, K.M. (2006) Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc Natl Acad Sci U S A*, **103**, 2677-2682.
171. Kha, D.T., Wang, G., Natrajan, N., Harrison, L. and Vasquez, K.M. (2010) Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. *Journal of molecular biology*, **398**, 471-480.
172. Jain, A., Bacolla, A., Del Mundo, I.M., Zhao, J., Wang, G. and Vasquez, K.M. (2013) DHX9 helicase is involved in preventing genomic instability

induced by alternatively structured DNA in human cells. *Nucleic Acids Res*, **41**, 10345-10357.
173. Kolb, J., Chuzhanova, N.A., Hogel, J., Vasquez, K.M., Cooper, D.N., Bacolla, A. and Kehrer-Sawatzki, H. (2009) Cruciform-forming inverted repeats appear to have mediated many of the microinversions that distinguish the human and chimpanzee genomes. *Chromosome Res*, **17**, 469-483.
174. Howard-Flanders, P., West, S.C. and Stasiak, A. (1984) Role of RecA protein spiral filaments in genetic recombination. *Nature*, **309**, 215-219.
175. Rao, B.J. and Radding, C.M. (1994) Formation of base triplets by non-Watson-Crick bonds mediates homologous recognition in RecA recombination filaments. *Proc Natl Acad Sci U S A*, **91**, 6161-6165.
176. Biet, E., Sun, J.S. and Dutreix, M. (2003) Stimulation of D-loop formation by polypurine/polypyrimidine sequences. *Nucleic Acids Res*, **31**, 1006-1012.
177. Rao, B.J., Dutreix, M. and Radding, C.M. (1991) Stable three-stranded DNA made by RecA protein. *Proc Natl Acad Sci U S A*, **88**, 2984-2988.
178. Zhurkin, V.B., Raghunathan, G., Ulyanov, N.B., Camerini-Otero, R.D. and Jernigan, R.L. (1994) A parallel DNA triplex as a model for the intermediate in homologous recombination. *J Mol Biol*, **239**, 181-200.
179. Collier, D.A., Griffin, J.A. and Wells, R.D. (1988) Non-B right-handed DNA conformations of homopurine.homopyrimidine sequences in the murine immunoglobulin C alpha switch region. *J Biol Chem*, **263**, 7397-7405.
180. Rao, B.S. (1994) Pausing of simian virus 40 DNA replication fork movement in vivo by (dG-dA)n.(dT-dC)n tracts. *Gene*, **140**, 233-237.
181. Krasilnikova, M.M. and Mirkin, S.M. (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol*, **24**, 2286-2295.
182. Wiener, F., Ohno, S., Babonits, M., Sumegi, J., Wirschubsky, Z., Klein, G., Mushinski, J.F. and Potter, M. (1984) Hemizygous interstitial deletion of chromosome 15 (band D) in three translocation-negative murine plasmacytomas. *Proc Natl Acad Sci U S A*, **81**, 1159-1163.
183. Joos, S., Haluska, F.G., Falk, M.H., Henglein, B., Hameister, H., Croce, C.M. and Bornkamm, G.W. (1992) Mapping chromosomal breakpoints of Burkitt's t(8;14) translocations far upstream of c-myc. *Cancer Res*, **52**, 6547-6552.
184. Kovalchuk, A.L., Muller, J.R. and Janz, S. (1997) Deletional remodeling of c-myc-deregulating chromosomal translocations. *Oncogene*, **15**, 2369-2377.
185. Akasaka, T., Akasaka, H., Ueda, C., Yonetani, N., Maesako, Y., Shimizu, A., Yamabe, H., Fukuhara, S., Uchiyama, T. and Ohno, H. (2000) Molecular and clinical features of non-Burkitt's, diffuse large-cell lymphoma of B-cell type associated with the c-MYC/immunoglobulin heavy-chain fusion gene. *J Clin Oncol*, **18**, 510-518.

186. Haluska, F.G., Tsujimoto, Y. and Croce, C.M. (1988) The t(8;14) breakpoint of the EW 36 undifferentiated lymphoma cell line lies 5' of MYC in a region prone to involvement in endemic Burkitt's lymphomas. *Nucleic Acids Res*, **16**, 2077-2085.
187. Saglio, G., Grazia Borrello, M., Guerrasio, A., Sozzi, G., Serra, A., di Celle, P.F., Foa, R., Ferrarini, M., Roncella, S., Borgna Pignatti, C. *et al.* (1993) Preferential clustering of chromosomal breakpoints in Burkitt's lymphomas and L3 type acute lymphoblastic leukemias with a t(8;14) translocation. *Genes Chromosomes Cancer*, **8**, 1-7.
188. Care, A., Cianetti, L., Giampaolo, A., Sposi, N.M., Zappavigna, V., Mavilio, F., Alimena, G., Amadori, S., Mandelli, F. and Peschle, C. (1986) Translocation of c-myc into the immunoglobulin heavy-chain locus in human acute B-cell leukemia. A molecular analysis. *EMBO J*, **5**, 905-911.
189. Wilda, M., Busch, K., Klose, I., Keller, T., Woessmann, W., Kreuder, J., Harbott, J. and Borkhardt, A. (2004) Level of MYC overexpression in pediatric Burkitt's lymphoma is strongly dependent on genomic breakpoint location within the MYC locus. *Genes Chromosomes Cancer*, **41**, 178-182.
190. Watnick, T.J., Piontek, K.B., Cordal, T.M., Weber, H., Gandolph, M.A., Qian, F., Lens, X.M., Neumann, H.P. and Germino, G.G. (1997) An unusual pattern of mutation in the duplicated portion of PKD1 is revealed by use of a novel strategy for mutation detection. *Hum Mol Genet*, **6**, 1473-1481.
191. Van Raay, T.J., Burn, T.C., Connors, T.D., Petry, L.R., Germino, G.G., Klinger, K.W. and Landes, G.M. (1996) A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. *Microb Comp Genomics*, **1**, 317-327.
192. Blaszak, R.T., Potaman, V., Sinden, R.R. and Bissler, J.J. (1999) DNA structural transitions within the PKD1 gene. *Nucleic Acids Res*, **27**, 2610-2617.
193. Akgun, E., Zahn, J., Baumes, S., Brown, G., Liang, F., Romanienko, P.J., Lewis, S. and Jasin, M. (1997) Palindrome resolution and recombination in the mammalian germ line. *Mol Cell Biol*, **17**, 5559-5570.
194. Nag, D.K. and Kurst, A. (1997) A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast Saccharomyces cerevisiae. *Genetics*, **146**, 835-847.
195. Tanaka, H., Tapscott, S.J., Trask, B.J. and Yao, M.C. (2002) Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. *Proc Natl Acad Sci U S A*, **99**, 8772-8777.
196. Zhou, Z.H., Akgun, E. and Jasin, M. (2001) Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. *Proc Natl Acad Sci U S A*, **98**, 8326-8333.

197. Cunningham, L.A., Cote, A.G., Cam-Ozdemir, C. and Lewis, S.M. (2003) Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism. *Mol Cell Biol*, **23**, 8740-8750.
198. Kurahashi, H., Shaikh, T.H., Hu, P., Roe, B.A., Emanuel, B.S. and Budarf, M.L. (2000) Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). *Hum Mol Genet*, **9**, 1665-1670.
199. Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M.G., Shanske, A., Goldberg, R. and Morrow, B.E. (2001) AT-rich palindromes mediate the constitutional t(11;22) translocation. *Am J Hum Genet*, **68**, 1-13.
200. Kurahashi, H. and Emanuel, B.S. (2001) Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum Mol Genet*, **10**, 2605-2617.
201. Kurahashi, H., Inagaki, H., Yamada, K., Ohye, T., Taniguchi, M., Emanuel, B.S. and Toda, T. (2004) Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. *J Biol Chem*, **279**, 35377-35383.
202. Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932-940.
203. Anvret, M., Ahlberg, G., Grandell, U., Hedberg, B., Johnson, K. and Edstrom, L. (1993) Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum Mol Genet*, **2**, 1397-1400.
204. Thornton, C.A., Johnson, K. and Moxley, R.T., 3rd. (1994) Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann Neurol*, **35**, 104-107.
205. Martorell, L., Martinez, J.M., Carey, N., Johnson, K. and Baiget, M. (1995) Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J Med Genet*, **32**, 593-596.
206. Wohrle, D., Kennerknecht, I., Wolf, M., Enders, H., Schwemmle, S. and Steinbach, P. (1995) Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation in vitro: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. *Hum Mol Genet*, **4**, 1147-1153.
207. Wong, L.J., Ashizawa, T., Monckton, D.G., Caskey, C.T. and Richards, C.S. (1995) Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet*, **56**, 114-122.
208. Zatz, M., Passos-Bueno, M.R., Cerqueira, A., Marie, S.K., Vainzof, M. and Pavanello, R.C. (1995) Analysis of the CTG repeat in skeletal muscle of young and adult myotonic dystrophy patients: when does the expansion occur? *Hum Mol Genet*, **4**, 401-406.
209. Martorell, L., Monckton, D.G., Gamez, J., Johnson, K.J., Gich, I., Lopez de Munain, A. and Baiget, M. (1998) Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum Mol Genet*, **7**, 307-312.

210. Kang, S., Jaworski, A., Ohshima, K. and Wells, R.D. (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet*, **10**, 213-218.
211. Miret, J.J., Pessoa-Brandao, L. and Lahue, R.S. (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **95**, 12438-12443.
212. Chong, S.S., McCall, A.E., Cota, J., Subramony, S.H., Orr, H.T., Hughes, M.R. and Zoghbi, H.Y. (1995) Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet*, **10**, 344-350.
213. Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J., Clarke, L.A. *et al.* (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet*, **6**, 409-414.
214. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. and Kanazawa, I. (1997) Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1. *Ann Neurol*, **41**, 505-511.
215. Ansved, T., Lundin, A. and Anvret, M. (1998) Larger CAG expansions in skeletal muscle compared with lymphocytes in Kennedy disease but not in Huntington disease. *Neurology*, **51**, 1442-1444.
216. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol Cancer Ther*, **1**, 565-573.
217. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, **99**, 11593-11598.
218. Wu, Y. and Brosh, R.M., Jr. (2010) G-quadruplex nucleic acids and human disease. *Febs J*, **277**, 3470-3488.
219. Juranek, S.A. and Paeschke, K. (2012) Cell cycle regulation of G-quadruplex DNA structures at telomeres. *Curr Pharm Des*, **18**, 1867-1872.
220. Capra, J.A., Paeschke, K., Singh, M. and Zakian, V.A. (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol*, **6**, e1000861.
221. Han, H., Hurley, L.H. and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res*, **27**, 537-542.
222. Wang, Q., Liu, J.Q., Chen, Z., Zheng, K.W., Chen, C.Y., Hao, Y.H. and Tan, Z. (2011) G-quadruplex formation at the 3' end of telomere DNA

inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res*, **39**, 6229-6237.
223. Paeschke, K., Capra, J.A. and Zakian, V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, **145**, 678-691.
224. Ogrunc, M., Becker, D.F., Ragsdale, S.W. and Sancar, A. (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol*, **180**, 5796-5798.
225. Wood, R.D. (1999) DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie*, **81**, 39-44.
226. Diderich, K., Alanazi, M. and Hoeijmakers, J.H. (2011) Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair (Amst)*, **10**, 772-780.
227. Missura, M., Buterin, T., Hindges, R., Hubscher, U., Kasparkova, J., Brabec, V. and Naegeli, H. (2001) Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair. *EMBO J*, **20**, 3554-3564.
228. Mitchell, D.L. and Nairn, R.S. (1989) The biology of the (6-4) photoproduct. *Photochem Photobiol*, **49**, 805-819.
229. Gunz, D., Hess, M.T. and Naegeli, H. (1996) Recognition of DNA adducts by human nucleotide excision repair. Evidence for a thermodynamic probing mechanism. *J Biol Chem*, **271**, 25089-25098.
230. Buschta-Hedayat, N., Buterin, T., Hess, M.T., Missura, M. and Naegeli, H. (1999) Recognition of nonhybridizing base pairs during nucleotide excision repair of DNA. *Proc Natl Acad Sci U S A*, **96**, 6090-6095.
231. Wang, G., Seidman, M.M. and Glazer, P.M. (1996) Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science*, **271**, 802-805.
232. Wang, G., Chen, Z., Zhang, S., Wilson, G.L. and Jing, K. (2001) Detection and determination of oligonucleotide triplex formation-mediated transcription-coupled DNA repair in HeLa nuclear extracts. *Nucleic Acids Res*, **29**, 1801-1807.
233. Vasquez, K.M., Christensen, J., Li, L., Finch, R.A. and Glazer, P.M. (2002) Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. *Proc Natl Acad Sci U S A*, **99**, 5848-5853.
234. Thoma, B.S., Wakasugi, M., Christensen, J., Reddy, M.C. and Vasquez, K.M. (2005) Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. *Nucleic Acids Res*, **33**, 2993-3001.
235. Christensen, L.A., Wang, H., Van Houten, B. and Vasquez, K.M. (2008) Efficient processing of TFO-directed psoralen DNA interstrand crosslinks by the UvrABC nuclease. *Nucleic Acids Res*, **36**, 7136-7145.

236. Parniewski, P., Bacolla, A., Jaworski, A. and Wells, R.D. (1999) Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli. *Nucleic Acids Res*, **27**, 616-623.
237. Oussatcheva, E.A., Hashem, V.I., Zou, Y., Sinden, R.R. and Potaman, V.N. (2001) Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in Escherichia coli. *J Biol Chem*, **276**, 30878-30884.
238. Lin, Y., Dion, V. and Wilson, J.H. (2006) Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol*, **13**, 179-180.
239. Lin, Y. and Wilson, J.H. (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol Cell Biol*, **27**, 6209-6217.
240. Szwarocka, S.T., Staczek, P. and Parniewski, P. (2007) Chromosomal model for analysis of a long CTG/CAG tract stability in wild-type Escherichia coli and its nucleotide excision repair mutants. *Can J Microbiol*, **53**, 860-868.
241. Lang, W.H., Coats, J.E., Majka, J., Hura, G.L., Lin, Y., Rasnik, I. and McMurray, C.T. (2011) Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. *Proc Natl Acad Sci U S A*, **108**, E837-844.
242. Iyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev*, **106**, 302-323.
243. Lange, S.S., Takata, K. and Wood, R.D. (2011) DNA polymerases and cancer. *Nat Rev Cancer*, **11**, 96-110.
244. Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem*, **65**, 101-133.
245. Lange, S.S., Mitchell, D.L. and Vasquez, K.M. (2008) High mobility group protein B1 enhances DNA repair and chromatin modification after DNA damage. *P Natl Acad Sci USA*, **105**, 10320-10325.
246. de Rosa, M., de Sanctis, D., Rosario, A.L., Archer, M., Rich, A., Athanasiadis, A. and Carrondo, M.A. (2010) Crystal structure of a junction between two Z-DNA helices. *Proc Natl Acad Sci U S A*, **107**, 9088-9092.
247. Taghian, D.G., Hough, H. and Nickoloff, J.A. (1998) Biased short tract repair of palindromic loop mismatches in mammalian cells. *Genetics*, **148**, 1257-1268.
248. Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M. and McMurray, C.T. (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol*, **12**, 663-670.

249. Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet*, **23**, 471-473.
250. van den Broek, W.J., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., te Riele, H., Groenen, P.J. and Wieringa, B. (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum Mol Genet*, **11**, 191-198.
251. Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P. and Monckton, D.G. (2004) Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum Mol Genet*, **13**, 1815-1825.
252. Wu, Q., Christensen, L.A., Legerski, R.J. and Vasquez, K.M. (2005) Mismatch repair participates in error-free processing of DNA interstrand crosslinks in human cells. *EMBO Rep*, **6**, 551-557.
253. Wu, Q. and Vasquez, K.M. (2008) Human MLH1 protein participates in genomic damage checkpoint signaling in response to DNA interstrand crosslinks, while MSH2 functions in DNA repair. *PLoS Genet*, **4**, e1000189.
254. Wang, G. and Vasquez, K.M. (2009) Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol Carcinog*, **48**, 286-298.
255. Schroth, G.P., Chou, P.J. and Ho, P.S. (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *The Journal of biological chemistry*, **267**, 11846-11855.
256. Vasquez, K.M. and Wang, G. (2013) The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutat Res*, **743-744**, 118-131.
257. Oh, D.B., Kim, Y.G. and Rich, A. (2002) Z-DNA-binding proteins can act as potent effectors of gene expression in vivo. *P Natl Acad Sci USA*, **99**, 16666-16671.
258. Wolfl, S., Wittig, B. and Rich, A. (1995) Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim Biophys Acta*, **1264**, 294-302.
259. Gessner, R.V., Frederick, C.A., Quigley, G.J., Rich, A. and Wang, A.H. (1989) The molecular structure of the left-handed Z-DNA double helix at 1.0-A atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J Biol Chem*, **264**, 7921-7935.
260. Schulz, V.P. and Zakian, V.A. (1994) The saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. *Cell*, **76**, 145-155.
261. Callahan, J.L., Andrews, K.J., Zakian, V.A. and Freudenreich, C.H. (2003) Mutations in yeast replication proteins that increase CAG/CTG expansions

also increase repeat fragility. *Molecular and cellular biology*, **23**, 7849-7860.
262. Wang, G., Zhao, J. and Vasquez, K.M. (2009) Methods to determine DNA structural alterations and genetic instability. *Methods*, **48**, 54-62.
263. Freudenreich, C.H., Kantrow, S.M. and Zakian, V.A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, **279**, 853-856.
264. Conde, J. and Fink, G.R. (1976) A mutant of Saccharomyces cerevisiae defective for nuclear fusion. *Proc Natl Acad Sci U S A*, **73**, 3651-3655.
265. Nilssontillgren, T., Petersen, J.G.L., Holmberg, S. and Kiellandbrandt, M.C. (1980) Transfer of Chromosome-Iii during Kar Mediated Cytoduction in Yeast. *Carlsberg Res Commun*, **45**, 113-117.
266. Zhu, X.D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J.H. and de Lange, T. (2003) ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol Cell*, **12**, 1489-1498.
267. Wu, Y., Zacal, N.J., Rainbow, A.J. and Zhu, X.D. (2007) XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. *DNA Repair (Amst)*, **6**, 157-166.
268. Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol*, **26**, 365-369.
269. Mukherjee, A. and Vasquez, K.M. (2016) HMGB1 interacts with XPA to facilitate the processing of DNA interstrand crosslinks in human cells. *Nucleic Acids Res*, **44**, 1151-1160.
270. Nelson, K.L., Becker, N.A., Pahwa, G.S., Hollingsworth, M.A. and Maher, L.J., 3rd. (1996) Potential for H-DNA in the human MUC1 mucin gene promoter. *J Biol Chem*, **271**, 18061-18067.
271. Boeke, J.D., LaCroute, F. and Fink, G.R. (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet*, **197**, 345-346.
272. Boeke, J.D., Trueheart, J., Natsoulis, G. and Fink, G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol*, **154**, 164-175.
273. Atomi, H., Imanaka, T. and Fukui, T. (2012) Overview of the genetic tools in the Archaea. *Front Microbiol*, **3**, 337.
274. Maruyama, A., Mimura, J., Harada, N. and Itoh, K. (2013) Nrf2 activation is associated with Z-DNA formation in the human HO-1 promoter. *Nucleic Acids Res*, **41**, 5223-5234.
275. Garner, M.M. and Felsenfeld, G. (1987) Effect of Z-DNA on nucleosome placement. *J Mol Biol*, **196**, 581-590.
276. Vasquez, K.M. and Wilson, J.H. (1998) Triplex-directed modification of genes and gene activity. *Trends Biochem Sci*, **23**, 4-9.

277. Blaho, J.A., Larson, J.E., McLean, M.J. and Wells, R.D. (1988) Multiple DNA secondary structures in perfect inverted repeat inserts in plasmids. Right-handed B-DNA, cruciforms, and left-handed Z-DNA. *J Biol Chem*, **263**, 14446-14455.
278. Ikegami, S., Taguchi, T., Ohashi, M., Oguro, M., Nagano, H. and Mano, Y. (1978) Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-alpha. *Nature*, **275**, 458-460.
279. Oguro, M., Suzuki-Hori, C., Nagano, H., Mano, Y. and Ikegami, S. (1979) The mode of inhibitory action by aphidicolin on eukaryotic DNA polymerase alpha. *Eur J Biochem*, **97**, 603-607.
280. Schiestl, R.H. and Prakash, S. (1988) RAD1, an excision repair gene of Saccharomyces cerevisiae, is also involved in recombination. *Mol Cell Biol*, **8**, 3619-3626.
281. Schiestl, R.H. and Prakash, S. (1990) RAD10, an excision repair gene of Saccharomyces cerevisiae, is involved in the RAD1 pathway of mitotic recombination. *Mol Cell Biol*, **10**, 2485-2491.
282. Saparbaev, M., Prakash, L. and Prakash, S. (1996) Requirement of mismatch repair genes MSH2 and MSH3 in the RAD1-RAD10 pathway of mitotic recombination in Saccharomyces cerevisiae. *Genetics*, **142**, 727-736.
283. Langston, L.D. and Symington, L.S. (2005) Opposing roles for DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 in yeast gene targeting. *EMBO J*, **24**, 2214-2223.
284. Fishman-Lobell, J. and Haber, J.E. (1992) Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. *Science*, **258**, 480-484.
285. Bardwell, A.J., Bardwell, L., Tomkinson, A.E. and Friedberg, E.C. (1994) Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science*, **265**, 2082-2085.
286. Ivanov, E.L. and Haber, J.E. (1995) RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in Saccharomyces cerevisiae. *Mol Cell Biol*, **15**, 2245-2251.
287. Sugawara, N., Paques, F., Colaiacovo, M. and Haber, J.E. (1997) Role of Saccharomyces cerevisiae Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proc Natl Acad Sci U S A*, **94**, 9214-9219.
288. Ciccia, A., McDonald, N. and West, S.C. (2008) Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem*, **77**, 259-287.
289. Paques, F. and Haber, J.E. (1997) Two pathways for removal of nonhomologous DNA ends during double-strand break repair in Saccharomyces cerevisiae. *Mol Cell Biol*, **17**, 6765-6771.

290. Evans, E., Sugawara, N., Haber, J.E. and Alani, E. (2000) The Saccharomyces cerevisiae Msh2 mismatch repair protein localizes to recombination intermediates in vivo. *Mol Cell*, **5**, 789-799.
291. Bertrand, P., Tishkoff, D.X., Filosi, N., Dasgupta, R. and Kolodner, R.D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. *Proc Natl Acad Sci U S A*, **95**, 14278-14283.
292. Kirkpatrick, D.T. and Petes, T.D. (1997) Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins. *Nature*, **387**, 929-931.
293. Kearney, H.M., Kirkpatrick, D.T., Gerton, J.L. and Petes, T.D. (2001) Meiotic recombination involving heterozygous large insertions in Saccharomyces cerevisiae: formation and repair of large, unpaired DNA loops. *Genetics*, **158**, 1457-1476.
294. Nicholson, A., Fabbri, R.M., Reeves, J.W. and Crouse, G.F. (2006) The effects of mismatch repair and RAD1 genes on interchromosomal crossover recombination in Saccharomyces cerevisiae. *Genetics*, **173**, 647-659.
295. Jensen, L.E., Jauert, P.A. and Kirkpatrick, D.T. (2005) The large loop repair and mismatch repair pathways of Saccharomyces cerevisiae act on distinct substrates during meiosis. *Genetics*, **170**, 1033-1043.
296. Lan, L., Hayashi, T., Rabeya, R.M., Nakajima, S., Kanno, S., Takao, M., Matsunaga, T., Yoshino, M., Ichikawa, M., Riele, H. *et al.* (2004) Functional and physical interactions between ERCC1 and MSH2 complexes for resistance to cis-diamminedichloroplatinum(II) in mammalian cells. *DNA Repair (Amst)*, **3**, 135-143.
297. Zhang, N., Liu, X., Li, L. and Legerski, R. (2007) Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. *DNA Repair (Amst)*, **6**, 1670-1678.
298. Zhang, N., Lu, X., Zhang, X., Peterson, C.A. and Legerski, R.J. (2002) hMutSbeta is required for the recognition and uncoupling of psoralen interstrand cross-links in vitro. *Mol Cell Biol*, **22**, 2388-2397.
299. Zhao, J., Bacolla, A., Wang, G. and Vasquez, K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci*, **67**, 43-62.
300. Wang, G. and Vasquez, K.M. (2014) Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst)*, **19**, 143-151.
301. Janz, S., Muller, J., Shaughnessy, J. and Potter, M. (1993) Detection of recombinations between c-myc and immunoglobulin switch alpha in murine plasma cell tumors and preneoplastic lesions by polymerase chain reaction. *Proc Natl Acad Sci U S A*, **90**, 7361-7365.

302. Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M. and Wood, R.D. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J*, **16**, 6559-6573.
303. Schwartz, E.K. and Heyer, W.D. (2011) Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes. *Chromosoma*, **120**, 109-127.
304. Hohl, M., Thorel, F., Clarkson, S.G. and Scharer, O.D. (2003) Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. *J Biol Chem*, **278**, 19500-19508.
305. Wu, X., Li, J., Li, X., Hsieh, C.L., Burgers, P.M. and Lieber, M.R. (1996) Processing of branched DNA intermediates by a complex of human FEN-1 and PCNA. *Nucleic Acids Res*, **24**, 2036-2043.
306. Balakrishnan, L. and Bambara, R.A. (2013) Flap endonuclease 1. *Annu Rev Biochem*, **82**, 119-138.
307. Zheng, L., Dai, H., Zhou, M., Li, M., Singh, P., Qiu, J., Tsark, W., Huang, Q., Kernstine, K., Zhang, X. *et al.* (2007) Fen1 mutations result in autoimmunity, chronic inflammation and cancers. *Nat Med*, **13**, 812-819.
308. Larsen, E., Gran, C., Saether, B.E., Seeberg, E. and Klungland, A. (2003) Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. *Mol Cell Biol*, **23**, 5346-5353.
309. Staresincic, L., Fagbemi, A.F., Enzlin, J.H., Gourdin, A.M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S.G., Vermeulen, W. and Scharer, O.D. (2009) Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J*, **28**, 1111-1120.
310. Jones, C.J., Cleaver, J.E. and Wood, R.D. (1992) Repair of damaged DNA by extracts from a xeroderma pigmentosum complementation group A revertant and expression of a protein absent in its parental cell line. *Nucleic Acids Res*, **20**, 991-995.
311. Zheng, H., Wang, X., Legerski, R.J., Glazer, P.M. and Li, L. (2006) Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. *DNA Repair (Amst)*, **5**, 566-574.
312. Umar, A., Koi, M., Risinger, J.I., Glaab, W.E., Tindall, K.R., Kolodner, R.D., Boland, C.R., Barrett, J.C. and Kunkel, T.A. (1997) Correction of hypermutability, N-methyl-N'-nitro-N-nitrosoguanidine resistance, and defective DNA mismatch repair by introducing chromosome 2 into human tumor cells with mutations in MSH2 and MSH6. *Cancer Res*, **57**, 3949-3955.
313. Aquilina, G., Ceccotti, S., Martinelli, S., Soddu, S., Crescenzi, M., Branch, P., Karran, P. and Bignami, M. (2000) Mismatch repair and p53 independently affect sensitivity to N-(2-chloroethyl)-N'-cyclohexyl-N-nitrosourea. *Clin Cancer Res*, **6**, 671-680.
314. Branch, P., Masson, M., Aquilina, G., Bignami, M. and Karran, P. (2000) Spontaneous development of drug resistance: mismatch repair and p53

defects in resistance to cisplatin in human tumor cells. *Oncogene*, **19**, 3138-3145.
315. Blasi, M.F., Ventura, I., Aquilina, G., Degan, P., Bertario, L., Bassi, C., Radice, P. and Bignami, M. (2006) A human cell-based assay to evaluate the effects of alterations in the MLH1 mismatch repair gene. *Cancer Res*, **66**, 9036-9044.
316. Seidman, M.M., Dixon, K., Razzaque, A., Zagursky, R.J. and Berman, M.L. (1985) A shuttle vector plasmid for studying carcinogen-induced point mutations in mammalian cells. *Gene*, **38**, 233-237.
317. Storici, F., Henneke, G., Ferrari, E., Gordenin, D.A., Hubscher, U. and Resnick, M.A. (2002) The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. *EMBO J*, **21**, 5930-5942.
318. Evans, E., Fellows, J., Coffer, A. and Wood, R.D. (1997) Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *EMBO J*, **16**, 625-638.
319. Kaushik, S., Kaushik, M., Svinarchuk, F., Malvy, C., Fermandjian, S. and Kukreti, S. (2011) Presence of divalent cation is not mandatory for the formation of intramolecular purine-motif triplex containing human c-jun protooncogene target. *Biochemistry*, **50**, 4132-4142.
320. Hohl, M., Dunand-Sauthier, I., Staresincic, L., Jaquier-Gubler, P., Thorel, F., Modesti, M., Clarkson, S.G. and Scharer, O.D. (2007) Domain swapping between FEN-1 and XPG defines regions in XPG that mediate nucleotide excision repair activity and substrate specificity. *Nucleic Acids Res*, **35**, 3053-3063.
321. Shen, B., Nolan, J.P., Sklar, L.A. and Park, M.S. (1997) Functional analysis of point mutations in human flap endonuclease-1 active site. *Nucleic Acids Res*, **25**, 3332-3338.
322. Boyer, A.S., Grgurevic, S., Cazaux, C. and Hoffmann, J.S. (2013) The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. *J Mol Biol*, **425**, 4767-4781.
323. Liu, Y., Nairn, R.S. and Vasquez, K.M. (2009) Targeted gene conversion induced by triplex-directed psoralen interstrand crosslinks in mammalian cells. *Nucleic Acids Res*, **37**, 6378-6388.
324. Christensen, L.A., Wang, H., Van Houten, B. and Vasquez, K.M. (2008) Efficient processing of TFO-directed psoralen DNA interstrand crosslinks by the UvrABC nuclease. *Nucleic Acids Res*, **36**, 7136-7145.
325. Richards, S., Liu, S.T., Majumdar, A., Liu, J.L., Nairn, R.S., Bernier, M., Maher, V. and Seidman, M.M. (2005) Triplex targeted genomic crosslinks enter separable deletion and base substitution pathways. *Nucleic Acids Res*, **33**, 5382-5393.
326. Barre, F.X., Asseline, U. and Harel-Bellan, A. (1999) Asymmetric recognition of psoralen interstrand crosslinks by the nucleotide excision repair and the error-prone repair pathways. *J Mol Biol*, **286**, 1379-1387.

327. Wood, R.D. (1996) DNA repair in eukaryotes. *Annu Rev Biochem*, **65**, 135-167.
328. Friedberg, E.C. (2003) DNA damage and repair. *Nature*, **421**, 436-440.
329. de Laat, W.L., Appeldoorn, E., Jaspers, N.G. and Hoeijmakers, J.H. (1998) DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem*, **273**, 7835-7842.
330. Fagbemi, A.F., Orelli, B. and Scharer, O.D. (2011) Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst)*, **10**, 722-729.
331. Vasquez, K.M. and Legerski, R.J. (2010) DNA interstrand crosslinks: repair, cell signaling, and therapeutic implications. *Environ Mol Mutagen*, **51**, 491-492.
332. Rahn, J.J., Adair, G.M. and Nairn, R.S. (2010) Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. *Environ Mol Mutagen*, **51**, 567-581.
333. Lee, S.K., Yu, S.L., Prakash, L. and Prakash, S. (2002) Requirement of yeast RAD2, a homolog of human XPG gene, for efficient RNA polymerase II transcription. implications for Cockayne syndrome. *Cell*, **109**, 823-834.
334. Al Mamun, A.A., Lombardo, M.J., Shee, C., Lisewski, A.M., Gonzalez, C., Lin, D., Nehring, R.B., Saint-Ruf, C., Gibson, J.L., Frisch, R.L. *et al.* (2012) Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science*, **338**, 1344-1348.
335. Umar, A. and Kunkel, T.A. (1996) DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur J Biochem*, **238**, 297-307.
336. Modrich, P. (2006) Mechanisms in eukaryotic mismatch repair. *J Biol Chem*, **281**, 30305-30309.
337. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, **7**, 335-346.
338. Mirkin, S.M. (2008) Discovery of alternative DNA structures: a heroic decade (1979-1989). *Front Biosci*, **13**, 1064-1071.
339. Harvey, S.C. (1983) DNA structural dynamics: longitudinal breathing as a possible mechanism for the B in equilibrium Z transition. *Nucleic Acids Res*, **11**, 4867-4878.
340. Sinden, R.R. (2005) Molecular biology: DNA twists and flips. *Nature*, **437**, 1097-1098.
341. Herbert, A., Lowenhaupt, K., Spitzner, J. and Rich, A. (1995) Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc Natl Acad Sci U S A*, **92**, 7550-7554.
342. Lafer, E.M., Sousa, R., Rosen, B., Hsu, A. and Rich, A. (1985) Isolation and characterization of Z-DNA binding proteins from wheat germ. *Biochemistry*, **24**, 5070-5076.
343. Kim, U., Wang, Y., Sanford, T., Zeng, Y. and Nishikura, K. (1994) Molecular cloning of cDNA for double-stranded RNA adenosine

deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci U S A*, **91**, 11457-11461.
344. O'Connell, M.A. and Keller, W. (1994) Purification and properties of double-stranded RNA-specific adenosine deaminase from calf thymus. *Proc Natl Acad Sci U S A*, **91**, 10596-10600.
345. Polson, A.G., Crain, P.F., Pomerantz, S.C., McCloskey, J.A. and Bass, B.L. (1991) The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry*, **30**, 11507-11514.
346. Bass, B.L. and Weintraub, H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, **55**, 1089-1098.
347. Bass, B.L. and Weintraub, H. (1987) A developmentally regulated activity that unwinds RNA duplexes. *Cell*, **48**, 607-613.
348. Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*, **71**, 817-846.
349. Wang, Q. (2011) RNA editing catalyzed by ADAR1 and its function in mammalian cells. *Biochemistry (Mosc)*, **76**, 900-911.
350. Herbert, A., Alfken, J., Kim, Y.G., Mian, I.S., Nishikura, K. and Rich, A. (1997) A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc Natl Acad Sci U S A*, **94**, 8421-8426.
351. Herbert, A., Schade, M., Lowenhaupt, K., Alfken, J., Schwartz, T., Shlyakhtenko, L.S., Lyubchenko, Y.L. and Rich, A. (1998) The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences. *Nucleic Acids Res*, **26**, 3486-3493.
352. George, C.X. and Samuel, C.E. (1999) Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. *Proc Natl Acad Sci U S A*, **96**, 4621-4626.
353. Bass, B.L., Nishikura, K., Keller, W., Seeburg, P.H., Emeson, R.B., O'Connell, M.A., Samuel, C.E. and Herbert, A. (1997) A standardized nomenclature for adenosine deaminases that act on RNA. *RNA*, **3**, 947-949.
354. Markle, D., Das, S., Ward, S.V. and Samuel, C.E. (2003) Functional analysis of the KCS-like element of the interferon-inducible RNA-specific adenosine deaminase ADAR1 promoter. *Gene*, **304**, 143-149.
355. Ward, S.V., Markle, D., Das, S. and Samuel, C.E. (2002) The promoter-proximal KCS element of the PKR kinase gene enhances transcription irrespective of orientation and position relative to the ISRE element and is functionally distinct from the KCS-like element of the ADAR deaminase Promoter. *J Interferon Cytokine Res*, **22**, 891-898.

356. Ha, S.C., Choi, J., Hwang, H.Y., Rich, A., Kim, Y.G. and Kim, K.K. (2009) The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1). *Nucleic Acids Res*, **37**, 629-637.
357. George, C.X., Gan, Z., Liu, Y. and Samuel, C.E. (2011) Adenosine deaminases acting on RNA, RNA editing, and interferon action. *J Interferon Cytokine Res*, **31**, 99-117.
358. Lai, F., Drakas, R. and Nishikura, K. (1995) Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. *The Journal of biological chemistry*, **270**, 17098-17105.
359. Xiao, W. and SpringerLink (Online service). *Methods in Molecular Biology, Methods and Protocols,*. 3rd ed, pp. X, 313 p. 357 illus., 327 illus. in color.
360. Chakraborty, P. and Grosse, F. (2010) WRN helicase unwinds Okazaki fragment-like hybrids in a reaction stimulated by the human DHX9 helicase. *Nucleic Acids Res*, **38**, 4722-4730.
361. Compton, S.A., Tolun, G., Kamath-Loeb, A.S., Loeb, L.A. and Griffith, J.D. (2008) The Werner syndrome protein binds replication fork and holliday junction DNAs as an oligomer. *J Biol Chem*, **283**, 24478-24483.
362. Kitao, H., Nanda, I., Sugino, R.P., Kinomura, A., Yamazoe, M., Arakawa, H., Schmid, M., Innan, H., Hiom, K. and Takata, M. (2011) FancJ/Brip1 helicase protects against genomic losses and gains in vertebrate cells. *Genes Cells*, **16**, 714-727.
363. Bacolla, A., Wang, G., Jain, A., Chuzhanova, N.A., Cer, R.Z., Collins, J.R., Cooper, D.N., Bohr, V.A. and Vasquez, K.M. (2011) Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of base substitution in human cells. *J Biol Chem*, **286**, 10017-10026.
364. Chu, W.K. and Hickson, I.D. (2009) RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer*, **9**, 644-654.
365. Cantor, S.B. and Guillemette, S. (2011) Hereditary breast cancer and the BRCA1-associated FANCJ/BACH1/BRIP1. *Future Oncol*, **7**, 253-261.
366. Brosh, R.M., Jr., Majumdar, A., Desai, S., Hickson, I.D., Bohr, V.A. and Seidman, M.M. (2001) Unwinding of a DNA triple helix by the Werner and Bloom syndrome helicases. *J Biol Chem*, **276**, 3024-3030.
367. Sommers, J.A., Rawtani, N., Gupta, R., Bugreev, D.V., Mazin, A.V., Cantor, S.B. and Brosh, R.M., Jr. (2009) FANCJ uses its motor ATPase to destabilize protein-DNA complexes, unwind triplexes, and inhibit RAD51 strand exchange. *J Biol Chem*, **284**, 7505-7517.
368. London, T.B., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J. and Hiom, K. (2008) FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J Biol Chem*, **283**, 36132-36139.
369. Fry, M. and Loeb, L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J Biol Chem*, **274**, 12797-12802.

370. Popuri, V., Bachrati, C.Z., Muzzolini, L., Mosedale, G., Costantini, S., Giacomini, E., Hickson, I.D. and Vindigni, A. (2008) The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J Biol Chem*, **283**, 17766-17776.
371. Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, **70**, 369-413.
372. Jaworski, A., Higgins, N.P., Wells, R.D. and Zacharias, W. (1991) Topoisomerase mutants and physiological conditions control supercoiling and Z-DNA formation in vivo. *J Biol Chem*, **266**, 2576-2581.
373. Spitzner, J.R., Chung, I.K. and Muller, M.T. (1990) Eukaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats. *Nucleic Acids Res*, **18**, 1-11.
374. Zacharias, W., Caserta, M., O'Connor, T.R., Larson, J.E. and Wells, R.D. (1988) Cytosine methylation as an effector of right-handed to left-handed DNA structural transitions. *Gene*, **74**, 221-224.
375. Fujii, S., Wang, A.H., van der Marel, G., van Boom, J.H. and Rich, A. (1982) Molecular structure of (m5 dC-dG)3: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res*, **10**, 7879-7892.
376. Behe, M. and Felsenfeld, G. (1981) Effects of methylation on a synthetic polynucleotide: the B--Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc Natl Acad Sci U S A*, **78**, 1619-1623.
377. Lee, J.S., Woodsworth, M.L., Latimer, L.J. and Morgan, A.R. (1984) Poly(pyrimidine) . poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res*, **12**, 6603-6614.
378. Xodo, L.E., Manzini, G., Quadrifoglio, F., van der Marel, G.A. and van Boom, J.H. (1991) Effect of 5-methylcytosine on the stability of triple-stranded DNA--a thermodynamic study. *Nucleic Acids Res*, **19**, 5625-5631.
379. Bacolla, A., Temiz, N.A., Yi, M., Ivanic, J., Cer, R.Z., Donohue, D.E., Ball, E.V., Mudunuri, U.S., Wang, G., Jain, A. *et al.* (2013) Guanine holes are prominent targets for mutation in cancer and inherited disease. *PLoS Genet*, **9**, e1003816.
380. Ogden, C.L., Carroll, M.D., Kit, B.K. and Flegal, K.M. (2014) Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama*, **311**, 806-814.
381. Lichtman, M.A. (2010) Obesity and the risk for a hematological malignancy: leukemia, lymphoma, or myeloma. *Oncologist*, **15**, 1083-1101.
382. Dobbins, M., Decorby, K. and Choi, B.C. (2013) The Association between Obesity and Cancer Risk: A Meta-Analysis of Observational Studies from 1985 to 2011. *ISRN Prev Med*, **2013**, 680536.

383. Calle, E.E., Rodriguez, C., Walker-Thurmond, K. and Thun, M.J. (2003) Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*, **348**, 1625-1638.
384. Kaidar-Person, O., Bar-Sela, G. and Person, B. (2011) The two major epidemics of the twenty-first century: obesity and cancer. *Obes Surg*, **21**, 1792-1797.
385. Calle, E.E. and Kaaks, R. (2004) Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*, **4**, 579-591.
386. Vucenik, I. and Stains, J.P. (2012) Obesity and cancer risk: evidence, mechanisms, and recommendations. *Ann N Y Acad Sci*, **1271**, 37-43.
387. Akasaka, T., Akasaka, H., Ueda, C., Yonetani, N., Maesako, Y., Shimizu, A., Yamabe, H., Fukuhara, S., Uchiyama, T. and Ohno, H. (2000) Molecular and clinical features of non-Burkitt's, diffuse large-cell lymphoma of B-cell type associated with the c-MYC/immunoglobulin heavy-chain fusion gene. *J Clin Oncol*, **18**, 510-518.
388. Damas, J., Carneiro, J., Goncalves, J., Stewart, J.B., Samuels, D.C., Amorim, A. and Pereira, F. (2012) Mitochondrial DNA deletions are associated with non-B DNA conformations. *Nucleic Acids Res*, **40**, 7606-7621.
389. Lee, J.E., Heo, J.I., Park, S.H., Kim, J.H., Kho, Y.J., Kang, H.J., Chung, H.Y., Yoon, J.L. and Lee, J.Y. (2011) Calorie restriction (CR) reduces age-dependent decline of non-homologous end joining (NHEJ) activity in rat tissues. *Exp Gerontol*, **46**, 891-896.
390. Sharma, S., Choudhary, B. and Raghavan, S.C. (2011) Efficiency of nonhomologous DNA end joining varies among somatic tissues, despite similarity in mechanism. *Cell Mol Life Sci*, **68**, 661-676.
391. ACS. (2016). American Cancer Society, http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/, pp. 1.
392. Benayoun, B.A., Pollina, E.A. and Brunet, A. (2015) Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol*, **16**, 593-610.
393. Pedersen, J.K., Engholm, G., Skytthe, A., Christensen, K. and Academy of Geriatric Cancer, R. (2016) Cancer and aging: Epidemiology and methodological challenges. *Acta Oncol*, **55 Suppl 1**, 7-12.
394. Campisi, J. (2005) Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell*, **120**, 513-522.
395. Maslov, A.Y. and Vijg, J. (2009) Genome instability, cancer and aging. *Biochimica et biophysica acta*, **1790**, 963-969.
396. Milholland, B., Auton, A., Suh, Y. and Vijg, J. (2015) Age-related somatic mutations in the cancer genome. *Oncotarget*, **6**, 24627-24635.

397. Sturm, A., Ivics, Z. and Vellai, T. (2015) The mechanism of ageing: primary role of transposable elements in genome disintegration. *Cell Mol Life Sci*, **72**, 1839-1847.
398. Siudeja, K., Nassari, S., Gervais, L., Skorski, P., Lameiras, S., Stolfa, D., Zande, M., Bernard, V., Rio Frio, T. and Bardin, A.J. (2015) Frequent Somatic Mutation in Adult Intestinal Stem Cells Drives Neoplasia and Genetic Mosaicism during Aging. *Cell Stem Cell*, **17**, 663-674.
399. Vyjayanti, V.N. and Rao, K.S. (2006) DNA double strand break repair in brain: reduced NHEJ activity in aging rat neurons. *Neurosci Lett*, **393**, 18-22.
400. Puthiyaveetil, A.G. and Caudell, D.L. (2013) Non homologous end joining-mediated DNA break repair is impaired in B lymphocytes of aging mice. *Mol Immunol*, **53**, 79-87.
401. Vaidya, A., Mao, Z., Tian, X., Spencer, B., Seluanov, A. and Gorbunova, V. (2014) Knock-in reporter mice demonstrate that DNA repair by non-homologous end joining declines with age. *PLoS genetics*, **10**, e1004511.
402. Seo, A.Y. and Leeuwenburgh, C. (2015) The Role of Genome Instability in Frailty: Mitochondria versus Nucleus. *Nestle Nutr Inst Workshop Ser*, **83**, 19-27.
403. Lee, S.Y., Lee, H., Kim, E.S., Park, S., Lee, J. and Ahn, B. (2015) WRN translocation from nucleolus to nucleoplasm is regulated by SIRT1 and required for DNA repair and the development of chemoresistance. *Mutat Res*, **774**, 40-48.
404. Nicolai, S., Rossi, A., Di Daniele, N., Melino, G., Annicchiarico-Petruzzelli, M. and Raschella, G. (2015) DNA repair and aging: the impact of the p53 family. *Aging (Albany NY)*, **7**, 1050-1065.
405. Kalfalah, F., Seggewiss, S., Walter, R., Tigges, J., Moreno-Villanueva, M., Burkle, A., Ohse, S., Busch, H., Boerries, M., Hildebrandt, B. *et al.* (2015) Structural chromosome abnormalities, increased DNA strand breaks and DNA strand break repair deficiency in dermal fibroblasts from old female human donors. *Aging (Albany NY)*, **7**, 110-122.
406. Vermeij, W.P., Hoeijmakers, J.H. and Pothof, J. (2016) Genome Integrity in Aging: Human Syndromes, Mouse Models, and Therapeutic Options. *Annu Rev Pharmacol Toxicol*, **56**, 427-445.
407. Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H. and McMurray, C.T. (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*, **447**, 447-452.
408. Hubert, L., Jr., Lin, Y., Dion, V. and Wilson, J.H. (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum Mol Genet*, **20**, 4822-4830.
409. Greenblatt, M.S., Bennett, W.P., Hollstein, M. and Harris, C.C. (1994) Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res*, **54**, 4855-4878.

410. Hainaut, P., Hernandez, T., Robinson, A., Rodriguez-Tome, P., Flores, T., Hollstein, M., Harris, C.C. and Montesano, R. (1998) IARC Database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools. *Nucleic Acids Res*, **26**, 205-213.
411. Prives, C. and Hall, P.A. (1999) The p53 pathway. *J Pathol*, **187**, 112-126.
412. Roy, R., Rodig, S., Bielenberg, D., Zurakowski, D. and Moses, M.A. (2011) ADAM12 transmembrane and secreted isoforms promote breast tumor growth: a distinct role for ADAM12-S protein in tumor metastasis. *J Biol Chem*, **286**, 20758-20768.
413. Roy, R., Wewer, U.M., Zurakowski, D., Pories, S.E. and Moses, M.A. (2004) ADAM 12 cleaves extracellular matrix proteins and correlates with cancer status and stage. *J Biol Chem*, **279**, 51323-51330.
414. Peduto, L., Reuter, V.E., Sehara-Fujisawa, A., Shaffer, D.R., Scher, H.I. and Blobel, C.P. (2006) ADAM12 is highly expressed in carcinoma-associated stroma and is required for mouse prostate tumor progression. *Oncogene*, **25**, 5462-5466.
415. Le Pabic, H., Bonnier, D., Wewer, U.M., Coutand, A., Musso, O., Baffet, G., Clement, B. and Theret, N. (2003) ADAM12 in human liver cancers: TGF-beta-regulated expression in stellate cells is associated with matrix remodeling. *Hepatology*, **37**, 1056-1066.
416. Frohlich, C., Albrechtsen, R., Dyrskjot, L., Rudkjaer, L., Orntoft, T.F. and Wewer, U.M. (2006) Molecular profiling of ADAM12 in human bladder cancer. *Clin Cancer Res*, **12**, 7359-7368.
417. Kodama, T., Ikeda, E., Okada, A., Ohtsuka, T., Shimoda, M., Shiomi, T., Yoshida, K., Nakada, M., Ohuchi, E. and Okada, Y. (2004) ADAM12 is selectively overexpressed in human glioblastomas and is associated with glioblastoma cell proliferation and shedding of heparin-binding epidermal growth factor. *Am J Pathol*, **165**, 1743-1753.
418. Tian, B.L., Wen, J.M., Zhang, M., Xie, D., Xu, R.B. and Luo, C.J. (2002) The expression of ADAM12 (meltrin alpha) in human giant cell tumours of bone. *Mol Pathol*, **55**, 394-397.
419. Pories, S.E., Zurakowski, D., Roy, R., Lamb, C.C., Raza, S., Exarhopoulos, A., Scheib, R.G., Schumer, S., Lenahan, C., Borges, V. *et al.* (2008) Urinary metalloproteinases: noninvasive biomarkers for breast cancer risk assessment. *Cancer Epidemiol Biomarkers Prev*, **17**, 1034-1042.
420. Kveiborg, M., Frohlich, C., Albrechtsen, R., Tischler, V., Dietrich, N., Holck, P., Kronqvist, P., Rank, F., Mercurio, A.M. and Wewer, U.M. (2005) A role for ADAM12 in breast tumor progression and stromal cell apoptosis. *Cancer Res*, **65**, 4754-4761.
421. Ray, B.K., Dhar, S., Henry, C., Rich, A. and Ray, A. (2013) Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-

12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family. *Cancer Res*, **73**, 736-744.
422. Membrino, A., Cogoi, S., Pedersen, E.B. and Xodo, L.E. (2011) G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One*, **6**, e24421.
423. Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. and Hurley, L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J Med Chem*, **40**, 2113-2116.
424. Sun, D., Guo, K. and Shin, Y.J. (2011) Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Res*, **39**, 1256-1265.
425. Heinrich, M.C., Corless, C.L., Demetri, G.D., Blanke, C.D., von Mehren, M., Joensuu, H., McGreevey, L.S., Chen, C.J., Van den Abbeele, A.D., Druker, B.J. *et al.* (2003) Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol*, **21**, 4342-4349.
426. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, **10**, 261-275.
427. Wang, K., Busker-Mannie, A.E., Hoeft, J., Vasquez, K., Miller, S.D., Melvold, R.W. and Waltenbaugh, C. (2001) Prolonged Hya-disparate skin graft survival in ethanol-consuming mice: correlation with impaired delayed hypersensitivity. *Alcohol Clin Exp Res*, **25**, 1542-1548.

Vita

Jennifer Andrea McKinney was born in Albuquerque, New Mexico. After graduating from Cibola High School in 1999, she began her undergraduate work at The University of New Mexico in Albuquerque, New Mexico. She completed her undergraduate degree at Texas State University in San Marcos, Texas, earning a Bachelor of Science in Biochemistry in 2006. While obtaining her degree, she was employed as a full-time clinical laboratory assistant and emergency room technician at Brackenridge Hospital in Austin, Texas. Jennifer continued her academic career by earning a Master of Science in Biochemistry from Texas State University in San Marcos, Texas in 2008. During the following year, she was employed as a lecturer at Texas State University in San Marcos, Texas. In January of 2010, she entered the Graduate School of Biomedical Sciences through the University of Texas MD Anderson Cancer Center at the Virginia Harris Cockrell Cancer Research Center at Science Park in Smithville, Texas. In February of 2011, she moved with her PI, Dr. Karen Vasquez, and entered the Graduate School at the University of Texas in Austin, Texas.

Permanent address (or email): 2424 Intrepid Dr., Buda, Texas, 78610

This dissertation was typed by Jennifer Andrea McKinney.